







RX	MEDLINE; 94112548.	FT	VARIANT	94	S -> A (IN INH-RESISTANT STRAIN NZ).
RA	BANERJEE A., DUBNAU E., QUENARD A., BALASUBRAMANIAN V., UM K.S., WILSON T., COLLINS D., DE LISLE G., JACOBS W.R. JR.,	SO	SEQUENCE	269 AA;	285/28 MW; 26DG955E CRC32;
RL	-!- CATALYTIC ACTIVITY: ACYL-[ACYL-CARRIER PROTEIN] + NAD(+) = TRANS-2,3-DIHYDROACYL-[ACYL-CARRIER PROTEIN] + NADH.		Query Match	39.5%	Score 342; DB 5; Length 269;
CC	-!- PATHWAY: SECOND REDUCTIVE STEP IN FATTY ACID BIOSYNTHESIS. THIS ISOZYME IS INVOLVED IN MYCOLIC ACID BIOSYNTHESIS. THIS		Best Local Similarity	33.5%	Pred. No. 7.37e-35;
CC	-!- INVOLVED IN THE RESISTANCE AGAINST THE ANTI-TUBERCULOSIS DRUGS ISONIAZID AND ETHIONAMIDE.		Matches	65;	Mismatches 93; Indels 19; Gaps 14;
CC	-!- SIMILARITY: SOME SIMILARITY TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES FAMILY (SDR).		Db	5 lgdkrlvsglttsiatihiarraqesqagqltv-gfdrl-iqrdr-lqitrdl-pakpl1 61	LENKTYVINGIANKRSIAGVAKYLDQGAKLVETYRRERSKELEKILEQLNQPEAHLY 63
DR	EMBL; U02330; G43704; -.		Qy	4 LENKTYVINGIANKRSIAGVAKYLDQGAKLVETYRRERSKELEKILEQLNQPEAHLY 63	62 eldvnneehlaslgrvtaeigagnklgdgvhsigfmpctgmginpfdapryadvtsgih 121
RW	OXIDOREDUCTASE; NAD; FATTY ACID BIOSYNTHESIS; ANTIBIOTIC RESISTANCE.		Db	44 QIDYQSDEVINGFEQIGKDVG --N-IDCVYSSIAFANMEDL-RGRSETSEGGFLAQD 119	QIDYQSDEVINGFEQIGKDVG --N-IDCVYSSIAFANMEDL-RGRSETSEGGFLAQD 119
NP_BIND	NP_BIND 136 165 NAD (POTENTIAL)		Qy	64 QIDYQSDEVINGFEQIGKDVG --N-IDCVYSSIAFANMEDL-RGRSETSEGGFLAQD 119	94 S -> A (IN INH-RESISTANT STRAIN NZ).
FT	FT VARIANT 94 94 SQUARED 651; SQUARED 651.		Db	122 isaysyassmakalpimpqggsvivgmdf-dpsramppnwmvtvakssalesvrfaareag 180	isaysyassmakalpimpqggsvivgmdf-dpsramppnwmvtvakssalesvrfaareag 180
FT	SEQUENCE 269 AA; 13ED04C CRC32;		Qy	120 ISSSLTIVTHEAKKLMPGGGSIVATVLLGGFAVQNYNMVGAVAKSLEANVKYLAIDLG 179	120 ISSSLTIVTHEAKKLMPGGGSIVATVLLGGFAVQNYNMVGAVAKSLEANVKYLAIDLG 179
SQ	Sequence 93; Conservative 59; Mismatches 92; Indels 19; Gaps 14;		Db	181 kygrsnlvaaapgirtlamsaivggalgeaqagiqileegdgrapigwmkda-pvax 240	kygrsnlvaaapgirtlamsaivggalgeaqagiqileegdgrapigwmkda-pvax 240
Db	5 legkrlytgittsdslafhiakvqaegaalvlvt-gfdrl-klyvriadiplkp-ap11 61		Qy	180 PDNTVRNNAISAGPIRLSAKG-VGG-F-N-T----ILKEI-EERAPLKRNV-DQVEYVK 228	5 legkrlytgittsdslafhiakvqaegaalvlvt-gfdrl-klyvriadiplkp-ap11 61
QY	4 LENKTYVINGIANKRSIAGVAKYLDQGAKLVETYRRERSKELEKILEQLNQPEAHLY 63		Db	241 tycallsdwlpartgdiyadgqta 266	4 LENKTYVINGIANKRSIAGVAKYLDQGAKLVETYRRERSKELEKILEQLNQPEAHLY 63
Db	62 eldvnneehlaslgrvtaeigagnklgdgvhsigfmpctgmginpfdapryadvtsgih 121		Qy	229 TAAYLSDLSGGVGENTHDGFHA 254	62 eldvnneehlaslgrvtaeigagnklgdgvhsigfmpctgmginpfdapryadvtsgih 121
QY	64 QIDYQSDEVINGFEQIGKDVG --N-IDCVYSSIAFANMEDL-RGRSETSEGGFLAQD 119		RESULT	8	SEQUENCE FROM N.A. (BAI1).
DB	122 isaysyassmakalpimpqggsvivgmdf-dpsramppnwmvtvakssalesvrfaareag 180		ID	BAI1_EUBSP	SEQUENCE FROM N.A. (BAI1).
QY	120 ISSSLTIVTHEAKKLMPGGGSIVATVLLGGFAVQNYNMVGAVAKSLEANVKYLAIDLG 179		AC	P07914	STANDARD; PRT; 249 AA.
DB	181 kygrsnlvaaapgirtlamsaivggalgeaqagiqileegdgrapigwmkda-pvax 240		DT	01-NOV-1996 (REL. 08, CREATED)	01-NOV-1996 (REL. 08, CREATED)
QY	180 PDNTVRNNAISAGPIRLSAKG-VGG-F-N-T----ILKEI-EERAPLKRNV-DQVEYVK 228		DT	01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)	01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
DB	241 tycallsdwlpartgdiyadgqta 266		DE	7-ALPHA-HYDROXYSTEROID DEHYDROGENASE (EC 1.1.1.159) (BILE ACID DE-7-DEHYDROXYLASE) (7-ALPHA-HSDH) (BILE ACID-INDUCIBLE PROTEIN).	7-ALPHA-HSDH (BILE ACID-INDUCIBLE PROTEIN).
QY	229 TAAYLSDLSGGVGENTHDGFHA 254		GN	BAI1 AND BAI2.	BAI1 AND BAI2.
OC	OC PROKARYOTA: FIRMICUTES; IRREGULAR ASPOROGENOUS RODS;		OS	BACTERIUM SP. (STRAIN VPI 12708).	BACTERIUM SP. (STRAIN VPI 12708).
RP	RP SEQUENCE FROM N.A. (BAI1).		RP	PROKARYOTA: FIRMICUTES; IRREGULAR ASPOROGENOUS RODS;	PROKARYOTA: FIRMICUTES; IRREGULAR ASPOROGENOUS RODS;
RX	RX MEDLINE: 90330548.		RP	SEQUENCE FROM N.A. (BAI1).	SEQUENCE FROM N.A. (BAI1).
RA	RA GOPAL SRIVASTAVA R., MAULJONE D.H., WHITE W.B., HYLEMON P.B.;		RP	RN [1]	RN [1]
RL	RL J. BACTERIOL. 172:4420-4426 (1990).		RN	RA COLEMAN J.P., WHITE W.B., LIJEWSKI M., HYLEMON P.B.;	RA COLEMAN J.P., WHITE W.B., LIJEWSKI M., HYLEMON P.B.;
CC	CC SEQUENCE FROM N.A. (BAI1).		RN	RA COLEMAN J.P., WHITE W.B., LIJEWSKI M., HYLEMON P.B.;	RA COLEMAN J.P., WHITE W.B., LIJEWSKI M., HYLEMON P.B.;
CC	CC MEDLINE: 8819793.		RN	RA COLEMAN J.P., WHITE W.B., LIJEWSKI M., HYLEMON P.B.;	RA COLEMAN J.P., WHITE W.B., LIJEWSKI M., HYLEMON P.B.;
CC	CC J. BACTERIOL. 170:2070-2077 (1988).		RN	RA COLEMAN J.P., WHITE W.B., LIJEWSKI M., HYLEMON P.B.;	RA COLEMAN J.P., WHITE W.B., LIJEWSKI M., HYLEMON P.B.;
CC	CC SEQUENCE OF 1-55 FROM N.A. (BAI1), AND SEQUENCE OF 1-33.		RP	RA COLEMAN J.P., WHITE W.B., LIJEWSKI M., HYLEMON P.B.;	RA COLEMAN J.P., WHITE W.B., LIJEWSKI M., HYLEMON P.B.;
CC	CC J. BACTERIOL. 169:1516-1521 (1987).		RP	RA COLEMAN J.P., WHITE W.B., LIJEWSKI M., HYLEMON P.B.;	RA COLEMAN J.P., WHITE W.B., LIJEWSKI M., HYLEMON P.B.;
CC	CC FUNCTION: 7-ALPHA-DEHYDROXYCHOLIC ACID, RESPECTIVELY, HIGHEST DEOXYCHOLIC ACID AND LITHOCHOLIC ACID.		RP	RA COLEMAN J.P., WHITE W.B., LIJEWSKI M., HYLEMON P.B.;	RA COLEMAN J.P., WHITE W.B., LIJEWSKI M., HYLEMON P.B.;
CC	CC AFFINITY WITH TAUROCHEMODEOXYCHOLIC ACID.		RP	RA COLEMAN J.P., WHITE W.B., LIJEWSKI M., HYLEMON P.B.;	RA COLEMAN J.P., WHITE W.B., LIJEWSKI M., HYLEMON P.B.;
CC	CC -!- CATALYTIC ACTIVITY: 3-ALPHA,7-ALPHA,12-ALPHA-TRIHYDROXY-5-BETA-CHOLANATE + NAD(+) = 3-ALPHA,7-ALPHA-DIHYDROXY-7-OXO-5-BETA-CHOLANATE + NADH.		RP	RA COLEMAN J.P., WHITE W.B., LIJEWSKI M., HYLEMON P.B.;	RA COLEMAN J.P., WHITE W.B., LIJEWSKI M., HYLEMON P.B.;
CC	CC -!- PATHWAY: BILE ACID METABOLISM.		RP	RA COLEMAN J.P., WHITE W.B., LIJEWSKI M., HYLEMON P.B.;	RA COLEMAN J.P., WHITE W.B., LIJEWSKI M., HYLEMON P.B.;
CC	CC -!- INDUCTION: PRESENCE OF C(24) BILE ACIDS CONTAINING A 7-ALPHA-HYDROXY GROUP.		RP	RA COLEMAN J.P., WHITE W.B., LIJEWSKI M., HYLEMON P.B.;	RA COLEMAN J.P., WHITE W.B., LIJEWSKI M., HYLEMON P.B.;
CC	CC -!- SIMILARITY: THERE ARE THREE GENES FOR BAI1A PROTEINS: BAI1A IS IDENTICAL TO BAI13 AND THERE IS 81% IDENTITY WITH BAI1A2.		RP	RA COLEMAN J.P., WHITE W.B., LIJEWSKI M., HYLEMON P.B.;	RA COLEMAN J.P., WHITE W.B., LIJEWSKI M., HYLEMON P.B.;
CC	CC -!- SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES FAMILY (SDR).		RP	RA COLEMAN J.P., WHITE W.B., LIJEWSKI M., HYLEMON P.B.;	RA COLEMAN J.P., WHITE W.B., LIJEWSKI M., HYLEMON P.B.;
DR	DR EMBL; M34658; GI48522;		RP	RA COLEMAN J.P., WHITE W.B., LIJEWSKI M., HYLEMON P.B.;	RA COLEMAN J.P., WHITE W.B., LIJEWSKI M., HYLEMON P.B.;
DR	DR EMBL; M19654; GI48516;		RP	RA COLEMAN J.P., WHITE W.B., LIJEWSKI M., HYLEMON P.B.;	RA COLEMAN J.P., WHITE W.B., LIJEWSKI M., HYLEMON P.B.;
DR	DR PIR; A28212; A26212.		RP	RA COLEMAN J.P., WHITE W.B., LIJEWSKI M., HYLEMON P.B.;	RA COLEMAN J.P., WHITE W.B., LIJEWSKI M., HYLEMON P.B.;
DR	DR PIR; A26938; A26938.		RP	RA COLEMAN J.P., WHITE W.B., LIJEWSKI M., HYLEMON P.B.;	RA COLEMAN J.P., WHITE W.B., LIJEWSKI M., HYLEMON P.B.;

DR PIR; B37762; B37762.  
 DR HSSP; P1992; 1HDC.  
 DR PROSTINE; PS0051; ADH\_SHORT.  
 KW OXIDOREDUCTASE; NAD; BILE ACID CATABOLISM; MULTIGENE FAMILY.  
 KW NP\_BIND 7 13  
 FT NP\_BIND 32 36  
 ACT SITE 157 162  
 FT CONFLICT 159 162 TSKA -> YOOG (IN REF. 2).  
 SQ SEQUENCE 249 AA; 26658 MW; 9E9D579 CRC32;

Query Match Score 229; DB 1; Length 249;  
 Best Local Similarity 23.7%; Pred. No. 4.16e-17;  
 Matches 61; Conservative 80; Mismatches 104; Indels 12; Gaps 12;

Db 1 mklvgdkitlgt-gr-rgqfaakdklfengatqvsifgetqeavdtaaqi-kelype 57  
 Qy 2 vnlf-lenttyvngiankrksrkelekkleqlnqpe 59

Db 58 evlgfpadlttsrdavnaavtgvaqkygrldminnagkt-mnsfsrvseedfkn-imdi 115  
 Qy 60 AHL-YQIDYQSDBEVINGEQIGKDGNIDCVYHSIAFANMEDGRSETSETREGFLIAQ 118

Db 116 nnygvngawswayqcmk-dakqgviintasvrtgiygsllsgiygptskagvgi1thglre 174  
 Qy 119 DISS-YSLTIVAHAEAKKLMPEGGSIIVATTYLGEGFAVQNYNMGYAKASLEANKYLAID 177

Db 175 iirknirvgyabgvyydmtmkgipp-e-1ledyltlpmkrmlkpeelamvylflasd1 232  
 Qy 178 LGPDNIRVNAISAGPRTLSAKGVGGENTILKEIEERAPLKRNVDOVEVSKTAAYLSDL 237

Db 233 asgitattisvdgayrp 249  
 Qy 238 SSGVGTGENITHVDSGFHA 254

RESULT 9 STANDARD; PRT; 249 AA.  
 ID BA72\_EUBSP  
 AC P19337;  
 DT 01-NOV-1990 (REL. 16, CREATED)  
 DT 01-NOV-1990 (REL. 16, LAST SEQUENCE UPDATE)  
 DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)  
 DE 7-ALPHA-HYDROXYSTEROL DEHYDROGENASE (EC 1.1.1.159) (BILE ACID  
 DE 7-DEHYDROXYLASE) (7-ALPHA-HSDH) (BILE ACID-INDUCIBLE PROTEIN).  
 GN BAI1A2.

OS EUBACTERIUM SP. (STRAIN VPI-12708),  
 OC PROKARYOTA; FIRMICTES; IRREGULAR ASPOROGENOUS RODS;  
 OC PROPRONIBACTERIAE.

[1] SEQUENCE FROM N.A.  
 MEDLINE; 91072233.  
 RA MALLONEE D.H., WHITE W.B., HYLEMON P.B.;  
 RL J. BACTERIOL.  
 RN SEQUENCE FROM N.A.  
 MEDLINE; 90330548.  
 RX GOPAL-SRIVASTAVA R., MALLONEE D.H., WHITE W.B., HYLEMON P.B.;  
 RA J. BACTERIOL. 177:4420-4426(1990).  
 RL J. BACTERIOL.  
 RN SEQUENCE FROM N.A.  
 RX SEQUENCE FROM N.A.  
 MEDLINE; 90264339.  
 RA MALLONEE D.H., WHITE W.B., HYLEMON P.B.;  
 RL J. BACTERIOL. 172:3400-3408(1990).

CC -!- PATHWAY: BILE ACID METABOLISM.  
 CC -!- INDUCTION: PRESENCE OF C(24) BILE ACIDS CONTAINING A  
 CC -!- 7-ALPHA-HYDROXY GROUP.  
 CC -!- SIMILARITY: THERE ARE THREE GENES FOR BAI1A PROTEINS; BAI1A IS IDENTICAL TO BAI1A3 AND THERE IS 81% IDENTITY WITH BAI1A2.  
 CC -!- SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES FAMILY (SDR).  
 CC EMBL; M36292; G148528;  
 DR EMBL; M22623; G290676;  
 DR EMBL; U54489; G1381567;  
 DR PIR; A31841; A31841.  
 DR PIR; E37844; E37844.  
 DR HSSP; P1992; 1HDC.  
 DR PROSITE; PS0061; ADH\_SHORT.  
 DR OXIDOREDUCTASE; NAD; BILE ACID CATABOLISM; MULTIGENE FAMILY.  
 KW NP\_BIND 7 13  
 FT NP\_BIND 32 36  
 ACT SITE 157 157 BY SIMILARITY.  
 SQ SEQUENCE 249 AA; 265318 MW; 453C0028 CRC32;

Query Match Score 225; DB 1; Length 249;  
 Best Local Similarity 23.7%; Pred. No. 1.65e-12;  
 Matches 61; Conservative 81; Mismatches 103; Indels 12; Gaps 12;

Db 1 mnlyqdktitlg-qt-rgqfaakifdingakvsiifgetqeavdtaaqi-kelype 57  
 Qy 2 LNL-ENKTYVNGIANKRKSRAFGVAKVLDQGAKL-VETYRKRSRKELEKULEQNQPE 59

Db 58 evlgfpadlttsrdavnaavtgvaqkygrldminnagkt-mnsfsrvseedfkn-imdi 115  
 Qy 60 AHL-YQIDYQSDBEVINGEQIGKDGNIDCVYHSIAFANMEDLRGPFSETSRGEFLIAQ 118

Db 116 nytgvngawswayqcmk-dakqgviintasvrtgiygsllsgiygptskagvgi1thglre 174  
 Qy 119 DISS-YSLTIVAHAEAKKLMPEGGSIIVATTYLGEGFAVQNYNMGYAKASLEANKYLAID 177

Db 175 iirknirvgyabgvyydmtmkgipp-e-1ledyltlpmkrmlkpeelamvylflasd1 232  
 Qy 178 LGPDNIRVNAISAGPRTLSAKGVGGENTILKEIEERAPLKRNVDOVEVSKTAAYLSDL 237

Db 233 asgitattisvdgayrp 249  
 Qy 238 SSGVGTGENITHVDSGFHA 254

RESULT 10 STANDARD; PRT; 244 AA.  
 ID AP27\_MOUSE  
 AC P08074;  
 DT 01-AUG-1988 (REL. 08, CREATED)  
 DT 01-AUG-1988 (REL. 08, LAST SEQUENCE UPDATE)  
 DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)  
 DE ADIPOCYTE P27 PROTEIN (AP27);  
 OS MUS MUSCULUS (MOUSE).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
 OC EUTHERIA; RODENTIA.  
 RN [1] SEQUENCE FROM N.A.  
 RP SEQUENCE FROM N.A.  
 PC STRAIN=CH3;  
 RX MEDLINE; 88273310.  
 RA NAVRE M., RINGOLD G.M.;  
 RL J. CELL BIOL. 107:219-286(1988).  
 CC -!- INDUCTION: THIS PROTEIN IS INDUCED BY GLUCOCORTICOIDS.  
 CC -!- SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES FAMILY (SDR).  
 CC EMBL; D26123; G599608;  
 DR EMBL; X74111; G50004;  
 DR PIR; A28053; A28053;  
 DR PIR; S03382; S03382.  
 DR PROSITE; PS0061; ADH\_SHORT.  
 KW OXIDOREDUCTASE.  
 FT NP\_BIND 11  
 FT ACT\_SITE 35  
 CC CATALYTIC ACTIVITY: 3-ALPHA-TAUROCHOLECYCLOLIC ACID.  
 CC -!- AFFINITY WITH TAUROCHOLECYCLOLIC ACID.  
 CC -!- CATALYTIC ACTIVITY: 3-ALPHA,12-ALPHA-TRIHYDROXY-7-OXO-5-BETA-CHOLANATE + NAD+ = 3-ALPHA,12-ALPHA-DIHYDROXY-7-OXO-5-BETA-CHOLANATE + NADH.  
 CC SEQUENCE 244 AA; 25958 MW; D8C37C9B CRC32;



NUCLEIC ACIDS RES.	14:7453-7472(1986).	Query Match Score 104%; Best Local Similarity 21.8%; Matches 55;	Score 183; Pred. No. 2.07e-10; Conservative 76; Mismatches 11;	DB 11; Gaps 7;
[2]	SEQUENCE FROM N.A.			
STRAIN:1021;				
RC FISHER R.F., SWANSON J.A., MULLIGAN J.T., LONG S.R.;				
CC GENETICS 117: 191-201(1987).				
-!- FUNCTION: PROPOSED TO MODIFY NOD FACTOR FATTY ACYL CHAIN				
-!- SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES				
FAMILY (SDR)				
EMBL: X04319; G46308; -.				
EMBL: Y00604; G46304; -.				
PIR: C24706; C24706.				
DR PIR: S07675; S07675.				
HSSP: P19999; 1HPC.				
DR PROSITE: PS00061; ADH_SHORT_NODULATION_OXIDOREDUCTASE_PLASMID_NAD.				
NP_BIND 35				
ACT_SITE 152				
SEQ_MATCH 245 AA;				
Best Local Similarity 24.1%;	Score 197; DB 6; Length 245;			
Matches 62; Conservative 73; Mismatches 108; Indels 14; Gaps 13;				
DDB 1 mfeIgtgrkalvgasg-aiggaiaarvlhagge-iv-ghlgcteik-letlataelgr-v 54		RESULT 15	PRT; 255 AA.	
: :  : :  : :  : :  : :  : :  : :  : :  : :  : :  : :  : :  : :		ID HDHA_ECOLI	STANDARD	
QY 1 MLNLENKTYVINGIANKRSIAFFGVAKYVLQDGAKLVEFTYKERSRKELEKLIQNOPEA 60		AC P2529;		
DDB 55 kifpanlanrdektalqrgaeadlegdilvnahmdapdw-dvlevn 112		DT 01-MAY-1992 (REL. 22, CREATED)		
: :  : :  : :  : :  : :  : :  : :  : :  : :  : :  : :  : :  : :		DT 01-OCT-1992 (REL. 22, LAST SEQUENCE UPDATE)		
QY 61 HLYQDVSDEEVINGFEGIDGVNIDQIVHSYAFAMEDLGRFSETSREGFLAQDI 120		DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)		
DDB 113 tamfrltreitq-qmirrnrgrinntsvalgqgnpgtqncaskagmifsfkslaqea 171		DE 7-ALPHA-HYDROXYSTEROID DEHYDROGENASE (EC 1.1.1.159) (7-ALPHA-HSDH).		
: :  : :  : :  : :  : :  : :  : :  : :  : :  : :  : :  : :  : :		GN HDHA OR HSDH.		
QY 121 SS-YSLTIVAEHKALKMPPEGGSIVATYLGGEFAVQNYNVMG/YAKASLEANVYKALDID 179		OS ESCHERICHIA COLI.		
DDB 172 trnitvncvapgf-e-sam-tdklnhkkkeklnvaipaihmrngtgtevasavaylasdh 228		OC PROKARYOTA; GRACILICUTES; SCOTOBACTERIA; FACULTATIVELY ANAEROBIC RODS; ENTEROBACTERIACEAE.		
: :  : :  : :  : :  : :  : :  : :  : :  : :  : :  : :  : :  : :		RN [1]	SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.	
QY 180 PDNRVNA-SAGPIRTLSAKGVGGFNTILKE-IEERAPLKRNVQDQEVEKTAAYLLSDLIS 238		RN [2]	RP STRAIN=HB101;	
DDB 229 ayvt9qtihvngmami 245		RN [3]	MDLINE: 91177603.	
QY 239 SGVIGENIHVDGFHAI 255		RN [4]	RA KURAZONO K.; TSURU D.; J. BACTERIOL. 173:2173-2179 (1991).	
DDB 14	RESULT	RP X-RAY CRYSTALLOGRAPHY (2.3 ANGSTROMS).		
CC ID YJGU_ECOLI STANDARD; PRT; 254 AA.		RX MEDLINE; 96244882.		
CC P3243; P7345; P7715; T7730(1996).		RA TANAKA N.; NONAKA T.; TANABE T.; YOSHIMOTO T.; TSURU D.; MITSUI Y.;		
CC BIOTECHNISTRY 35:7715-7730(1996).		RL DT 01-FEB-1995 (REL. 31, CREATED)		
CC -!- FUNCTION: 7-ALPHA-DEHYDROXYATION OF CHOLIC ACID, YIELDING		CC DT 01-FEB-1995 (REL. 31, LAST SEQUENCE UPDATE)		
CC DHOXYCHOLIC ACID AND LITHOCHOLIC ACID, RESPECTIVELY. HIGHEST		CC DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)		
CC AFFINITY WITH TAUROCHOLEDOXYCHOLIC ACID.		CC DT DE HYPOTHETICAL OXIDOREDUCTASE IN PEPA-GNTV INTERGENIC REGION		
CC -!- CATALYTIC ACTIVITY: 3-ALPHA,7-ALPHA,12-ALPHA-TRIHYDROXY-5-BETA-		CC DE HYPOTHETICAL OXIDOREDUCTASE IN PEPA-GNTV INTERGENIC REGION (EC 1.1.1.159) (7-ALPHA-HSDH).		
CC CHOLANATE + NAD(+) = 3-ALPHA,12-ALPHA-DIHYDROXY-7-OXO-5-BETA-		CC CC CHOLANATE + NADH.		
CC OS ESCHERICHIA COLI.		CC -!- SUBUNIT: HOMOTETRAMER.		
CC PROKARYOTA; GRACILICUTES; SCOTOBACTERIA; FACULTATIVELY ANAEROBIC RODS; ENTEROBACTERIACEAE.		CC -!- SUBUNIT: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES FAMILY (SDR).		
CC MEDLINE: 95134365.		CC DR EMBL: D10497; G216571; -.		
CC BURLAND V.D., PLUNKETT G. III, SOFIA H.J., DANIELS D.L., BLATTNER F.R.;		CC DR PIR: J70951; JT0951.		
CC NUCLEIC ACIDS RES. 23:2105-2119(1995).		CC DR HSSP; P19992; 1HDC.		
CC -!- SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES FAMILY (SDR).		CC DR ECOGENE; EG10425; HDHA.		
CC EMBL: U14003; G57108; -.		CC DR PROSITE: PS00061; ADH_SHORT_OXIDOREDUCTASE; NAD; BETA-CATENIN.		
CC ECOGENE; EG12540; YJGU.		CC KW KW HYPOTHETICAL PROTEIN; OXIDOREDUCTASE.		
CC PROSITE: PS00061; ADH_SHORT_OXIDOREDUCTASE.		CC FT NP_BIND 13 37 NAD OR NADP (BY SIMILARITY).		
CC FT ACT_SITE 158 158 BY SIMILARITY.		CC FT NP_BIND 18 24 NAD.		
CC SQ 27563 NW; 908F3466 CRC32;		CC FT ACT_SITE 159 159 NAD.		
CC SEQUENCE 254 AA; 26778 NW; 443F6382 CRC32;		CC SQ SEQUENCE 255 AA; 26778 NW; 443F6382 CRC32;		
CC		Query Match Score 10.2%; Best Local Similarity 25.7%; Matches 65; Conservative 74; Mismatches 101; Indels 13; Gaps 11;		

Fri Dec 5 12:31:32 1997

US-08-790-043A-1.rsp

Page 8

b 7 1rlqgkcaitg -aga-gigkeiaitfataragassvvsdinadaanhhvvdeiaggll-gqaf 63  
b 1 : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |  
y 2 LNLENKTTIVIMG.TANKRSIAFGYKVLDOLGAKVFTYKERSKELEKLLEQINOPPEAH 61

b 64 acrdctsegelsaladfa-is-k lgkyrdilvnnaggggpkp&ompmdfrr -ayelnv- 119  
b 1 : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |  
y 62 LRQDVQSDEEV - INGFFQIKGVNGIDGVYHISIAFANMEDLRGRFSETSERFQFLAQD 119

b 120 fsstflsqlyapemek - ngggyrltitsmaenkniuntstysskaashlvnrmfdl 177  
o 1 : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |  
y 120 ISSYSLT - IVAHZAKLIMPEGGSSVATYLGFFAVONYNYGVAKSLEANYKYLSDL 178

b 178 geknirringiaagailtdalksvit - peiegkmnlqhtpirrlrgqpdianalflcspaa 236  
o 1 : | : | : | : | : | : | : | : | : | : | : | : | : | : |  
y 179 GDPNIRNAISAPIRTLSAKGYGFNTILKETEAPLKRNYDQEVGKTAAYLLSDLS 238

b 237 swsgqiltsvsg 249  
o 1 : | : | : | : | : | : | : | : | : | : | : | : |

Job completed: Thu Dec 4 16:04:55 1997



**RESULT** 2  
**ID** R66293 standard; Protein; 269 AA.  
**AC** R66293;  
**DT** 03-JUL-1995 (first entry)  
**DE** M. smegmatis Inha.  
**RW** Isoniazid; isonicotinic acid hydrazide; INH; inhA gene; vaccine.  
**OS** Mycobacterium smegmatis.  
**PN** WO94/26765-A.  
**PD** 24-NOV-1994.  
**PR** 13-MAY-1994; U05398.  
**PR** 13-MAY-1993; NZ-24/620.  
**PR** 14-MAY-1993; US-052409.  
**PR** 31-MAR-1994; US-221742.  
**PA** (AGRE-) AGRESEARCH.  
**PA** (BANE-) BANERJEE A.  
**PA** (COLL-) COLLINS D.  
**PA** (DLIS-) DE LISLE G W.  
**PA** (JACO-) JACOBS W R.  
**PA** (WILS-) WILSON T M.  
**PA** (AGRE-) AGRESEARCH.  
**PA** (COLL-) COLLINS D.  
**PA** (YESH) UNIV YESHIVA EINSTEIN COLLEGE.  
**PI** Banerjee A, Collins DM, De LISLE GW, Jacobs WR;  
**PI** Wilson TM, Collins D;  
**WPI** 95-006366/01.

**PT** Gene target for isonicotinic acid hydrazide - used to develop prods for diagnosis, treatment, prevention and studies involving mycobacterial infections.

**PS** Disclosure; Fig. 4; 76pp; English.

**CC** Genes from Mycobacterium smegmatis, Mycobacterium tuberculosis and Mycobacterium bovis that encode inhA, the target of action for isoniazid, were identified, isolated and cloned. Sequences for the 3 genes are given in Q78913-15, and encoded amino acids in R66289-91. Mutant genes have been used in recombinant vaccine development.

**DR** WO94/26765-A, Collins D, De LISLE GW, Jacobs WR, Wilson TM; DR; WPI; 95-006651/01.

**DR** N-PDB; Q75517.

**PT** Polynucleotide(s) determining mycobacterial resistance to isoniazid - useful in diagnosis, treatment and prevention of mycobacterial infection, e.g. tuberculosis.

**PS** Disclosure; Fig. 4A-4B; 104pp; English.

**CC** The gene from Mycobacterium smegmatis, encoding Inha (R66293), the target of action for isoniazid, was identified, isolated, cloned and sequenced (Q75517). Mutant inhA genes have been used for recombinant vaccine development.

**Sequence** 269 AA;

**Query Match** 20.3%;  
**Best Local Similarity** 35.0%;  
**Matches** 92; **Conservative** 60; **Mismatches** 92; **Indels** 19; **Gaps** 14;

**Db** 5 legkrilvtgtitdssiafhiaqeaqeagelvl-t-gfdl-k-lvkriadr1pkp-ap1l 61

**Qy** 4 LENKTYVIMGANKRTRIAFGVAKVLDOLAKLVFYRKRSRKELEKLQLNQEAHLY 63

**Db** 62 eldvqeehhlstladritaegegnkidgvhaigfmqsgmdinpfdrapdqvrdvskgh 121

**Qy** 64 QDVSQDVEEINGFEQIGKVG--N-IDGVYHSLAFANMEDL-RGRFSETSRGFLLAQD 119

**Db** 122 isaystsaslakavlpimppggivgmdf-dptrrampnwtvaksalesvrfrvareag 180

**Qy** 120 ISSSVTIVHEAKLMPSEGSSTIVATTYLOGEFAVQNINVMGAKASLANYKLALDG 179

**Db** 181 kvgvrsnlvaagpirtlamsaiygalldeaggmgmlleegwdqraplgwnmkdpptvak 240

**Qy** 180 PDNRVNAISAGPRTLSAKG-VGG-F-N-T----ILKEI-BERAPLKRVN-DQVEVGK 228

**Db** 241 tvcallsdwlpattgtvイヤdg 263

**Qy** 229 TAAYLSDLSGTGENIHVDSG 251

**RESULT** 4  
**ID** R66290 standard; Protein; 269 AA.

**AC** R66290;  
**DT** 03-JUL-1995 (first entry)

**DE** M. tuberculosis inhA gene; isoniazid; isonicotinic acid hydrazide; INH; inhA gene; vaccine; mycobacteria; ds.

**OS** Mycobacterium tuberculosis.

**PN** WO94/26312-A.

**PD** 24-NOV-1994.

**AC** R66289; U05344.

**DT** 12-MAY-1994; U05344.

**PR** 13-MAY-1993; N-24/7620.

**PR** 14-MAY-1993; US-062409.

**PR** 31-MAR-1994; US-021742.

**PA** (BANE-) BANERJEE A.

**PA** (COLL-) COLLINS D M.

**PA** (DLIS-) DE LISLE G W.

**PA** (JACO-) JACOBS W R.

**PA** (WILS-) WILSON T M.

**PA** (AGRE-) AGRESEARCH.

**PA** (COLL-) COLLINS D.

**PA** (YESH) UNIV YESHIVA EINSTEIN COLLEGE.

**PI** Banerjee A, Collins DM, De LISLE GW, Jacobs WR;

**PI** Wilson TM, Collins D;

**PA** (BANE/) BANERJEE A.  
**PA** (COLL/) COLLINS D M.  
**PA** (DLIS/) DE LISLE G W.  
**PA** (JACO/) JACOBS W R.  
**PA** (WILS/) WILSON T M.  
**PA** (AGRE-) AGRESEARCH.  
**PA** (COLL/) COLLINS D.  
**PA** (YESH) UNIV YESHIVA EINSTEIN COLLEGE.  
**PI** Banerjee A, Collins DM, De LISLE GW, Jacobs WR;  
**PI** Wilson TM, Collins D;  
**WPI** 95-006366/01.

**PT** Gene target for isonicotinic acid hydrazide - used to develop prods for diagnosis, treatment, prevention and studies involving mycobacterial infections.

**PS** Disclosure; Fig. 4; 76pp; English.

**CC** Genes from Mycobacterium smegmatis, Mycobacterium tuberculosis and Mycobacterium bovis that encode inhA, the target of action for isoniazid, were identified, isolated and cloned. Sequences for the 3 genes are given in Q78913-15, and encoded amino acids in R66289-91. Mutant genes have been used in recombinant vaccine development.

**SQ** Sequence 269 AA;

**Query Match** 20.3%;  
**Best Local Similarity** 35.0%;  
**Matches** 92; **Conservative** 60; **Mismatches** 92; **Indels** 19; **Gaps** 14;

**Db** 5 legkrilvtgtitdssiafhiaqeaqeagelvl-t-gfdl-k-lvkriadr1pkp-ap1l 61

**Qy** 4 LENKTYVIMGANKRTRIAFGVAKVLDOLAKLVFYRKRSRKELEKLQLNQEAHLY 63

**Db** 62 eldvqeehhlstladritaegegnkidgvhaigfmqsgmdinpfdrapdqvrdvskgh 121

**Qy** 64 QDVSQDVEEINGFEQIGKVG--N-IDGVYHSLAFANMEDL-RGRFSETSRGFLLAQD 119

**Db** 122 isaystsaslakavlpimppggivgmdf-dptrrampnwtvaksalesvrfrvareag 180

**Qy** 120 ISSSVTIVHEAKLMPSEGSSTIVATTYLOGEFAVQNINVMGAKASLANYKLALDG 179

**Db** 181 kvgvrsnlvaagpirtlamsaiygalldeaggmgmlleegwdqraplgwnmkdpptvak 240

**Qy** 180 PDNRVNAISAGPRTLSAKG-VGG-F-N-T----ILKEI-BERAPLKRVN-DQVEVGK 228

**Db** 241 tvcallsdwlpattgtvイヤdg 263

**Qy** 229 TAAYLSDLSGTGENIHVDSG 251

**RESULT** 3  
**ID** R66289 standard; Protein; 269 AA.  
**AC** R66289;  
**DT** 03-JUL-1995 (first entry)  
**DE** M. smegmatis Inha.  
**RW** Isoniazid; isonicotinic acid hydrazide; INH; inhA gene; mycobacteria; vaccine.  
**OS** Mycobacterium smegmatis.  
**PN** WO94/26312-A.

**PD** 24-NOV-1994; U05344.

**PR** 12-MAY-1994; U05344.

**PR** 13-MAY-1993; N-24/7620.

**PR** 14-MAY-1993; US-062409.

**PR** 31-MAR-1994; US-021742.

**PA** (BANE-) BANERJEE A.

**PA** (COLL-) COLLINS D M.

**PA** (DLIS-) DE LISLE G W.

**PA** (JACO-) JACOBS W R.

**PA** (WILS-) WILSON T M.

**PA** (AGRE-) AGRESEARCH.

**PA** (COLL-) COLLINS D.

**PA** (YESH) UNIV YESHIVA EINSTEIN COLLEGE.

**PI** Banerjee A, Collins DM, De LISLE GW, Jacobs WR;

**PI** Wilson TM, Collins D;

WPI: 95-006366/01.  
 DR N-PSDB; 078914.  
 PT Gene target for isonicotinic acid hydrazide - used to develop  
 PT products for diagnosis, treatment, prevention and studies involving  
 PT mycobacterial infections  
 Disclosure; Fig. 4; 76pp; English.  
 PS Genes from Mycobacterium smegmatis, Mycobacterium tuberculosis and  
 CC Mycobacterium bovis that encode Inha, the target of action for  
 CC isoniazid, were identified, isolated and cloned. Sequences of  
 CC amino acids in Q8913\_15, and encoded amino acids in  
 CC R66289\_91. Mutant genes have been used in recombinant vaccine  
 development. 269 AA;

Query Match 19.4%; Score 340; DB 12; Length 269;  
 Best Local Similarity 32.7%; Pred. No. 8 65e-16; Gaps 14;  
 Matches 87; Conservative 87; Mismatches 68; Indels 19; Gaps 14;

Db 5 1dgkrilvsgitdssiafhiaarvagegqaglvlt-gfdrlr1-igritdr1-pakapl1 61

QY 4 LENKTYVIMGIANKRSTAFGAKLVFTYKERSKELELLEQNLQPEAHLY 63

Db 62 elivqneehlaslagrvteiagagnkldgryvhsgmpqtgmqinppffadpyadvskgh 121

QY 64 QIDYQSDEVINGFEQIGKDVG-N-IDGVYHSIAFANMEDL-RGRSETSEGFLLAQD 119

Db 122 isaygmdfoprallpimpqggsvmgdf-dpsrampaywmtvaksalesvrfrvareag 180

QY 120 ISSSLTIVAHAEKKLMPGGSVIATVILGEFAVQNYNNNGVAKASLEANYKYLALDLG 179

Db 181 kygrsnylygagpirtlamsaivgalgeagadiqliegwqrapigwnmkdatpvk 240

QY 180 PDNIRNATISAGPRTLSKG-VGG-F-NT---I-LKEI-EERAPLKRVN-DQVEVK 228

Db 241 tvcallsdwlpattdgdiyadggah 266

QY 229 TAAVLLSDLSSGTGENIHDSGPHA 254

RESULT 5. ID R66292 standard; Protein: 231 AA.

AC R66292; (First entry)

DT 03-JUL-1995

DE Mycobacterium bovis Inha.

KW Isoniazid; isonicotinic acid hydrazide; INH; InhA gene; pss5 gene;

KW vaccine; mycobacteria.

OS Mycobacterium bovis.

PA WO9426312-A.

PA (COLL/) COLLINS D.M.

PA (DISL/) DE LISLE G.W.

PA (JACO/) JACOB W.R.

PA (WILS/) WILSON T.M.

PA (WILS/) WILSON T.M.

PA (AGRE-) AGRESEARCH.

PA (COLL/) COLLINS D.

PA (YESH) UNIV YESHIVA EINSTEIN COLLEGE.

PA (BANE/) BANERJEE A., De LISLE GW, Jacobs WR;

PA (WILS/) WILSON TM, Collins D;

PI Wilson TM, Collins D;

DR 95-00636/01.

DR Q79915.

PT Gene target for isonicotinic acid hydrazide - used to develop  
 PT mycobacterial infections

PT Disclosure; Fig. 9A-9D; 76pp; English.

PT A cosmid library from M. bovis G4/100 was prepared in shuttle

PT vector pYUB18 and transformed into Mycobacterium smegmatis ms2155.

PT The smallest plasmid obtained which conferred resistance to Inha,

PT the target of action for isoniazid, was designated pS5. The

PT sequence of pss5 was determined (Q78915), revealing 2 large open

CC reading frames, ORF2 being the InhA gene encoding the protein given

CC in R66292.

SQ Sequence 231 AA;

Query Match 18.1%; Score 317; DB 12; Length 231;

Best Local Similarity 32.2%; Mismatches 77; Indels 9; Gaps 8;

Matches 66; Conservative 53; Mismatches 77; Indels 9;

PT 5 1dgkrilvsgitdssiafhiaarvagegqaglvlt-gfdr1r1-igritdr1-pakapl1 61

QY 4 LENKTYVIMGIANKRSTAFGAKLVFTYKERSKELELLEQNLQPEAHLY 63

Db 62 eldvneehlaslagrvteiagagnkldgryvhsgmpqtgmqinppffadpyadvskgh 121

QY 64 QIDYQSDEVINGFEQIGKDVG-N-IDGVYHSIAFANMEDL-RGRSETSEGFLLAQD 119

**RESULT** 7  
 ID R63900 standard; Protein; 231 AA.  
 AC R63900;  
 DT 03-JUL-1995 (first entry)  
 DE M. bovis InhA.  
 KW Isoniazid; isonicotinic acid hydrazide; INH; inhA gene; pS5; vaccine.  
 OS Mycobacterium bovis G4/100.  
 PN WO92118765-A.  
 PD 24-NOV-1994.  
 PT 13-MAY-1993; U05398.  
 14-MAY-1993; US-227620.  
 14-MAY-1993; US-062409.  
 31-MAR-1994; US-221742.  
 PA (AGRE-) AGRESEARCH.  
 PA (BANE/) BANERJEE.  
 PA (COLL/) COLLINS D.  
 PA (JRCO/) JACOBS W R.  
 PA (YESH/) UNIT YESHIVA EINSTEIN COLLEGE.  
 PA (WILS/) WILSON T M.  
 PI Banerjee A, Collins D, De Lisle GW, Jacobs WR, Wilson TM;  
 DR WPI; 95-006691/01.  
 PR • Q75519.  
 PT Polynucleotide(s) determining mycobacterial resistance to  
 PR isoniazid - useful in diagnosis, treatment and prevention of  
 PR mycobacterial infection, e.g. tuberculosis.  
 PS Disclosure: Page 21; 104pp; English.  
 CC A cosmid library of Mycobacterium bovis G4/100 was made in vector  
 PR pUB18 and transferred into Mycobacterium smegmatis mc2155. The  
 CC smallest plasmid obtained which conferred resistance to isoniazid resistance  
 CC phenotype on M. smegmatis was designated pS5. Sequencing of pS5  
 CC revealed 2 open reading frames, ORF2 encoded InhA (R6389),  
 CC the target of action for isoniazid. Mutants of the InhA gene  
 CC have been used for recombinant vaccine development.  
 SQ Sequence 231 AA;

Query Match 18.1%  
 Best Local Similarity 32.8%;  
 Matches 66; Conservative 53; Mismatches 77; Indels 9; Gaps 8;

5 ldgkrilvsgitissiafhiarrvageqgqavifl-9fdrlr-igritdrll-pakap1 61  
 Qy 4 LENKTYVINGTIANRSTIAFGVARYLDQGAKLVFTYKERSKKELVILEQINQPAHLY 63

Db 62 eldgneehlasigrvteaggnkldgyvhraigfmppqtgmngbfadpyadyskgih 121  
 Qy 64 QIDYQSDDEEVINGEQIICKDVG-N-IDGVVHSIAFANMEDI-LRGFSSETREGFLLAQD 119

Db 122 isaysasmakallplmmppggsvymdf-dpsrampaynmtvaksalesvnrvareag 180  
 Qy 120 ISSYSLTVAAHEKKLMPEGGSVATYLGGEFAYQVNNTMGVAKASLEANVKYLADLG 179

Db 181 kygyrsnlvaaqpirtlamsamgg 205  
 Qy 180 PDNIRVNAISAGPIRTLSAKGV-GG 203

Query Match 9.9%  
 Best Local Similarity 26.0%;  
 Matches 66; Conservative 73; Mismatches 100; Indels 15; Gaps 13;

Db 7 lrldgkaaitg-ag-a-gigkeiaittatgsavvsvdsinadaanhhvdeiqqlg-qgaf 63  
 Qy 2 LNLENKYVINGTIANRSTIAFGVARYLDQGAKLVFTYRERSRKELEKILEQNLQPEAH 61  
 Db 64 acrcdtsegealsaladfa-isk-lqkvdl-vnkpvdrgpkp-fdmpmdfr-rayelnv 119  
 Qy 62 LYQIDYQSDDEEV-INGFQIGKDVGND-VYHSIAFAMMEDLRGRFSETSREGFLLAQ 118  
 Db 120 -sfphlsqvapemek--rggylvlttsmaenknimtsyasskskaashlvrrnafd 176  
 Qy 119 DISSYSLT-TVAEAKKLMPEGGSIVATTYLGGEFAVQXNTMGVAKASLEANVKYLAD 177  
 Db 177 lgekirvnyriapqgqlltadalksvit-peieqknlightprrrlgqpqdianasalfcspa 235  
 Qy 178 LGPDNTRVNISAQGPIRTLSAKGVGGENTILKETEERAPLKRNVDOVEVKPAAYLLSDL 237  
 Db 236 aswvsgqqltvsgg 249  
 Qy 238 SSGVGTGENITHVDSG 251

RESULT 9  
 ID W02111 standard; Protein; 256 AA.  
 AC W02111;  
 DT 06-MAR-1997 (first entry)  
 DE Gluconate; NADP+-5'-oxidoreductase.  
 KW Gluconate; NADP; oxidoreductase; Gluconobacter oxydans;  
 OS Gluconobacter oxydans.  
 PD 14-AUG-1996.  
 PR 07-FEB-1995; 101776.  
 PA (RHEI-) RHEIN BIOTECH GES NEUE BIOTECHNOLOGISCHE.  
 PI Bringer-Meyer S, Hollenberg CP, Klasen R, Sahn H;  
 DR WPI; 95-364397-37.  
 N-PDB: T 36145.

PT 5-keto-9gluconate prodn. by increasing expression of gluconate;  
 PT NADP+-5'-oxido:reductase gene - esp by increasing copy no. in  
 Gluconobacter, used as intermediate for ascorbic and tartaric acids  
 Claim 8; Page 10-12; 15pp; German.  
 CC The Gluconobacter oxydans gene may be used to transform cells,  
 CC to produce higher levels of gluconate;NADP+-5'-oxidoreductase.  
 CC Tartaric acid can now be produced without the difficult  
 CC purification involved in prepn. from tartar.  
 SQ Sequence 256 AA;

RESULT 8  
 ID R22993 standard; Protein; 255 AA.  
 AC R22993;  
 DT 19-AUG-1992 (first entry)  
 DE 7-alpha-hydroxy steroid dehydrogenase.  
 KW Biochemical reagent.  
 OS Escherichia coli.

Query Match	9.8%; Score 171; DB 20; Length 256;	DB	133 keadv-eammktidaugtividrvnnnagitr-dtlilmrkksqwd-vidlnlnlgvfclt 189
Best Local Similarity	24.0%; Pred. No. 8.41e-04;	QY	69 SDEEVINGFEQIQKDV-GNIDGVYHSTAFAANMEDLRGRFSTISRGFLAQDISSYSLRI 127
Matches	59; Conservative 74; Mismatches 100; Indels 13; Gaps 13;	DB	190 qa-atkimmkkkrgriniavsvgligniganyaaakagrivfskttaaregasrnin 248
Db	15 lvgtasrglghtlarklarygaevvngrrnaidsaqsfea-eglkastavfdvtqd 73	QY	128 VAHEAKKLMP-E-GGSIVATTYLGEFAVQNTNMGYAKASLEANTKYLAIDLGPDPNIRVN 186
QY	12 MGIANKRISIAGVAKYLDOLGAKLYTYKERSRKELEKLELQEINOPSAHLQYDQSDE 71	DB	249 vvcpgfiiasdmak-lge-d-mekkilgtiplgryggpedavgreflalspaasyitq 305
Db	74 avidgyaaierdmgpidlinnagiqrraplee-fsrkwdw-lmstavnva-vf-fvgqa 129	QY	187 AISAPPIRPL-SAKGVGGFNTILKEIBERAPLKRNYDQVEVGKTAAYL-LSDLSSECVTGE 244
QY	72 ETINGFFQIGRDVGNDGTHSIAFANMDLRGRFSETSERGELLAQDISSTSLSLTVAHE 131	DB	306 aftidggi 313
Db	130 varhmiprgkvknicsyselarpgiaptakgavknltkmatqwgriqlingla 189	QY	245 NIHDSGF 252
QY	132 -ARKLMPEG-GSIVATTYLGEFAVQNTNMGYAKASLEANTKYLAIDLGPDPNIRVNAIS 189	DB	133 keadv-eammktidaugtividrvnnnagitr-dtlilmrkksqwd-vidlnlnlgvfclt 189
Db	190 pgfyatemterlvad-eeftdwlckrtptagrgyeevl-vg-aavflssrasffvngqvl 246	QY	69 SDEEVINGFEQIQKDV-GNIDGVYHSTAFAANMEDLRGRFSTISRGFLAQDISSYSLRI 127
Db	190 AGSPRT-LSAKGVGGFTNLREKIEERAPL-KRN-VDQEVEGKTAAYL-LSDLSSECVTGENI 246	DT	13-APR-1996 (first entry)
Db	247 mvdggi 252	DE	Rape seed beta-ketoadcy1-ACP-ketoreductase;
QY	247 HVDSGF 252	KW	Rape; seed; beta-ketoacyl-ACP-ketoreductase; beta-ketoreductase;
		KW	plasmid pors10.1; cDNA library; embryo; Escherichia coli; vector;
		KW	plastid; stroma; transit peptide; cassette; antisense; oilseed;
		KW	transgenic plant; crop improvement; lipid; metabolic engineering;
		KW	polymer; rapeseed oil; Brassica napus.
		OS	Brassica napus.
		FH	New isolated rape beta-ketoreductase DNA - used to develop plants
		FT	PT with lower or higher oil contents or with altered oil compsn.
		AC	Key
		AC	Location/Qualifiers
		AC	R89323 standard; Protein: 315 AA.
		AC	R89323; FT 1.56
		AC	/note= "Transit peptide"
		AC	FT 15; 29pp; English.
		AC	PN W09602652-2.
		AC	PD 01-FEB-1996.
		AC	PD 01-FEB-1996; WO9602652-2.
		AC	PF 17-JUL-1995; G01678.
		AC	PF 20-JUL-1994; GB-014622.
		AC	PA (ZENE ) ZENECA LTD.
		AC	PA Chase D, Elborough K, Fentem PA, Slabas AR, White A;
		AC	PI Chase D, Elborough K, Fentem PA, Slabas AR, White A;
		AC	DR WPI: 96-105914/11.
		AC	DR N-PSDB: Q99304.
		AC	PT New isolated rape beta-ketoreductase DNA - used to develop plants
		AC	PT with lower or higher oil contents or with altered oil compsn.
		AC	PS Claim 1; Page 15; 29pp; English.
		AC	CC The sequence corresponds to a rape seed beta-ketoreductase encoded by a cDNA insert in plasmid pors10.1 in Escherichia coli XL1-Blue.
		AC	CC A plastid stroma targeting transit peptide is present. DNA encoding the protein may be inserted in a vector or expression cassette in sense or antisense orientation for expression in an oilseed plant, e.g. for production of transgenic rape plants with low or modified oil content, diversion of metabolism to alternative storage compounds, e.g. starch, protein or engineered polymers, or production of plants with enhanced oil content. The DNA may also be used as a probe to obtain similar genes from other plants. The transit peptide may be used to direct other proteins to seed plastids.
		AC	CC Sequence 315 AA;
		AC	CC Best Local Similarity 23.4%; Pred. No. 2.09e-02; Matches 58; Conservative 67; Mismatches 110; Indels 13; Gaps 13;
		AC	CC DB 74 vvvvtgasrgkalslgkagckvlnvnyarsakeaeveksqieaqy-gqaitfggds 132
		AC	CC QY 10 VIMGIANKRSIAFGVAKVLDQLGAKLYTFY-RKERSRKELKEKLQLNPEAHLYQIDVQ 68
		AC	CC DB 133 keadv-eammktidaugtividrvnnnagitr-dtlilmrkksqwd-vidlnlnlgvfclt 189
		AC	CC QY 69 SDEEVINGFEQIQKDV-GNIDGVYHSTAFAANMEDLRGRFSTISRGFLAQDISSYSLRI 127
		AC	CC DB 190 qa-atkimmkkkrgriniavsvgligniganyaaakagrivfskttaaregasrnin 248
		AC	CC QY 128 VAHEAKKLMP-E-GGSIVATTYLGEFAVQNTNMGYAKASLEANTKYLAIDLGPDPNIRVN 186
		AC	CC DB 249 vvcpgfiiasdmak-lge-d-mekkilgtiplgryggpedavgveflalspaasyitq 305

QY 187 AISAGPIRT-LSAKGVGFNTILKEEERAPLKRNVTDQVEVGKTAAYL-LSDLSSGVTE 244  
 DB 306 affidggi 313  
 QY 245 NIHDSGF 252

**RESULT 12**  
 ID R10974 standard; Protein; 246 AA.  
 AC R10974; 91-051341/07.  
 DT 17-APR-1991 (first entry)  
 DE Acetoacetyl CoA reductase enzyme.  
 KW Polyester biopolymers; polyhydroxybutyrate; polyhydroxy alkanoate;  
 PT beta-ketothiolase; acetoacetyl CoA reductase.  
 KW Alcaligenes eutrophus.  
 OS WO9100517-A.  
 PN PP 10-JUL-1990; U038451.  
 PP 10-JUL-1989; US-378155.  
 PP 10-JUL-1991; U038451.  
 PP 10-JUL-1991; U038451.

This Alcaligenes eutrophus acetyl CoA reductase is an enzyme which is essential to the biosynthesis of polyhydroxyalcanoate (PHB). The gene encoding this is contained in plasmid clone pAE73, downstream from the thiolase gene. The use of recombinant methods for producing such enzymes, required for polyester biopolymer synthesis, allows for the control and modification of the synthesis process. See also WO9100517-A and Q10502-03. Sequence 246 AA.

Query Match Score 140; DB 2; Length 246;

Best Local Similarity 22.2%; Pred. No. 9.35e-02; Gaps 7;  
 Matches 41; Conservative 59; Mismatches 77; Indels 8; Gaps 7;

Db 66 dstktakdkvksrgedvlnnagitrdrvfr-kmtradvda-vldtnlts-lfnvtk 121  
 QY 71 EEWINGFEQKGDVKDGNIDGVHSTAFANMEDLRGREFSETSRREGFLAQDISYSLTIVAH 130

Db 122 qvdgmadrwgwgrivnissngqfqgtntystakaghfmlaqevtkgvtvty 181  
 QY 131 EAKKLMP-E-G-GSIVATTYLGGEFAVQVNLYMGYAKASLEANKYLAIDLGPDNIRVNAI 188

Db 182 spgyiatdmvkairq-d-vldkivatipkvrlglpeetasicawisseesgfsftgadfs1 239  
 QY 189 SAGPRTLSAKGVGCFNTILKEEBERAPLKRNVTDQVEVGKTAAYLLSDSSGVGENIHV 248

Db 240 nglh 244  
 QY 249 DSGFH 253

Query Match Score 140; DB 6; Length 246;

Best Local Similarity 22.2%; Pred. No. 9.35e-02; Gaps 7;  
 Matches 41; Conservative 59; Mismatches 77; Indels 8; Gaps 7;

Db 66 dstktakdkvksrgedvlnnagitrdrvfr-kmtradvda-vldtnlts-lfnvtk 121  
 QY 71 EEWINGFEQKGDVKDGNIDGVHSTAFANMEDLRGREFSETSRREGFLAQDISYSLTIVAH 130

Db 122 qvdgmadrwgwgrivnissngqfqgtntystakaghfmlaqevtkgvtvty 181  
 QY 131 EAKKLMP-E-G-GSIVATTYLGGEFAVQVNLYMGYAKASLEANKYLAIDLGPDNIRVNAI 188

Db 182 spgyiatdmvkairq-d-vldkivatipkvrlglpeetasicawisseesgfsftgadfs1 239  
 QY 189 SAGPRTLSAKGVGCFNTILKEEBERAPLKRNVTDQVEVGKTAAYLLSDSSGVGENIHV 248

Db 240 nglh 244  
 QY 249 DSGFH 253

RESULT 14

ID P94157 standard; protein; 246 AA.

AC P94157;

DT 12-JUN-1990 (first entry)

DE Acetyl-CoA reductase.

KW Polyester biopolymers; acetyl-CoA reductase; polyhydroxybutyrate; ss

OS Alcaligenes eutrophus H16.

PN WO890202-A.

PD 12-JUN-1989.

PR 29-JUN-1988; U02227.

PF 29-JUN-1987; US-0677695.

(MASI) Massachusetts Institute of Technology.

PA Peoples OP, Sinskey AJ;

DR WPI: 89-039655/05.

N-PSDB; N91209.

PT Constructing new polyester biopolymers - using genes encoding

beta-ketothiolases(s); acetoacetyl-CoA reductase(s).

PS Disclosure; p; English.

CC The enzyme is used to study the polyhydroxybutyrate (PHB) biosynthetic pathway. It catalyses the reduction of acetoacetyl CoA to form

CC D(-)-beta-hydroxybutyrate-CoA, the substrate for PHB synthetase, and can

CC be used to control biopolymer syntheses and produce new biopolymers.

CC The sequence was deduced from the DNA sequence.

CC See also P94154-P94156.

SEQ Sequence 246 AA;

DR WPI: 89-039655/05.

PT

PS

CC

CC

CC

CC

CC

CC

CC

CC

CC

Query Match Score 140; DB 1; Length 246;

Best Local Similarity 22.2%; Pred. No. 9.35e-02; Gaps 7;  
 Matches 41; Conservative 59; Mismatches 77; Indels 8; Gaps 7;

Db 66 dstktakdkvksrgedvlnnagitrdrvfr-kmtradvda-vldtnlts-lfnvtk 121  
 QY 71 EEWINGFEQKGDVKDGNIDGVHSTAFANMEDLRGREFSETSRREGFLAQDISYSLTIVAH 130

Query Match Score 13

ID R32192 standard; protein; 246 AA.

AC R32192;

DT 30-MAY-1993 (first entry)

DE Sequence encoded by the acetoacetyl-CoA reductase (phbB) gene of the

DE polyhydroxybutyrate (PHB) operon

KW Operon; polyhydroxyalcanoate; acetoacetyl-CoA reductase.

OS Alcaligenes eutrophus.

PN WO9302187-A.

PD 13-JUL-1992; U05786.

PF 19-JUL-1991; US-32243.

PA (UNMA) UNIV MADISON JAMES.

(UNMS) UNIV MICHIGAN STATE.

Db 122 qvidqmadrgwqrivnissvngqkgffgqtnystakaghftmalaevatkgvtvnty 181  
 Qy 131 EAKKLMP-E-G-SIVATTYLGKFDEFAVQNYNMGVARASLEANKVYLALDGPDNIRVNAI 188  
 Db 182 spgyiatdmvkairq-d-vldkivatipvkrlpeiasicawlsseesfstgadfs1 239  
 Qy 189 SAGPIRTLSAKGFGENTILKEEERAPLKRNYDQEVEGKTAAYLLSDLSSGTGENIHV 248  
 Db 240 nggih 244  
 Qy 249 DSGFH 253

## RESULT 15

TD R71324; standard; Protein; 329 AA.

AC R71324; (first entry)

DT 21-Oct-1995

DE Acetyl-CoA-reductase.

Acetyl-CoA-reductase; transgenic plant; poly-beta-hydroxalkanoate; poly-beta hydroxybutyrate; biodegradable thermoplastic.

Alcaligenes eutrophus.

OS WO950372 A.

PN 23-FEB-1995.

PD 17-AUG-1994; US09265.

PR 17-AUG-1993; US108193.

PR 06-JUN-1994; US25457.

PA (UNMS ) UNIV MICHIGAN STATE.

PI Nawrath C, Poirier Y, Somerville CR;

DR WPI: 95-098770/13.

DR N-PSDB: 081642.

PR Transgenic plant material with plastid(s) contg. the enzymes for synthesis of poly:hydroxy alkanoate(s) - express

PR poly:hydroxy butyrate and have good growth and seed formation.

PS Claim 2: Page 60-61; 88pp; English.

CC The acetyl-CoA-reductase gene (phbB), from A. eutrophus is cloned under the control of an Arabidopsis thaliana seed storage protein

CC promoter for plastid tissue-specific gene expression in a transgenic plant, when expressed with the 3-ketothiolase

CC (phbA) and poly-beta-hydroxalkanoate synthase (phbC) genes, a

CC poly-beta-hydroxalkanoate (PHB), specifically poly-beta-

CC hydroxybutyrate (PHB), is expressed in the transgenic plant

CC (preferably a Brassica e.g. rape). PHB and related PHAs are biodegradable thermoplastics with many useful applications.

SQ Sequence 329 AA;

Query Match 8.0%; Score 140; DB 13; Length 329;

Best Local Similarity 22.2%; Pred. No. 9.5e-02; Mismatches 7;

Matches 41; Conservative 59; Indels 8; Gaps 7;

Db 149 dscktafkvksseydvlunagnirdrdvir-kmrraduva-vidnlitv-lfnvtk 204

Qy 71 EEVINGFBIGKDGVNIDGVH3IAFAMMEDLRGRFFETSRGFFLAQDISSYSLTIVAH 130

Db 205 qvidqmadqgrivnissvngqkgffgqtnystakaghftmalaevatkgvtvnty 264

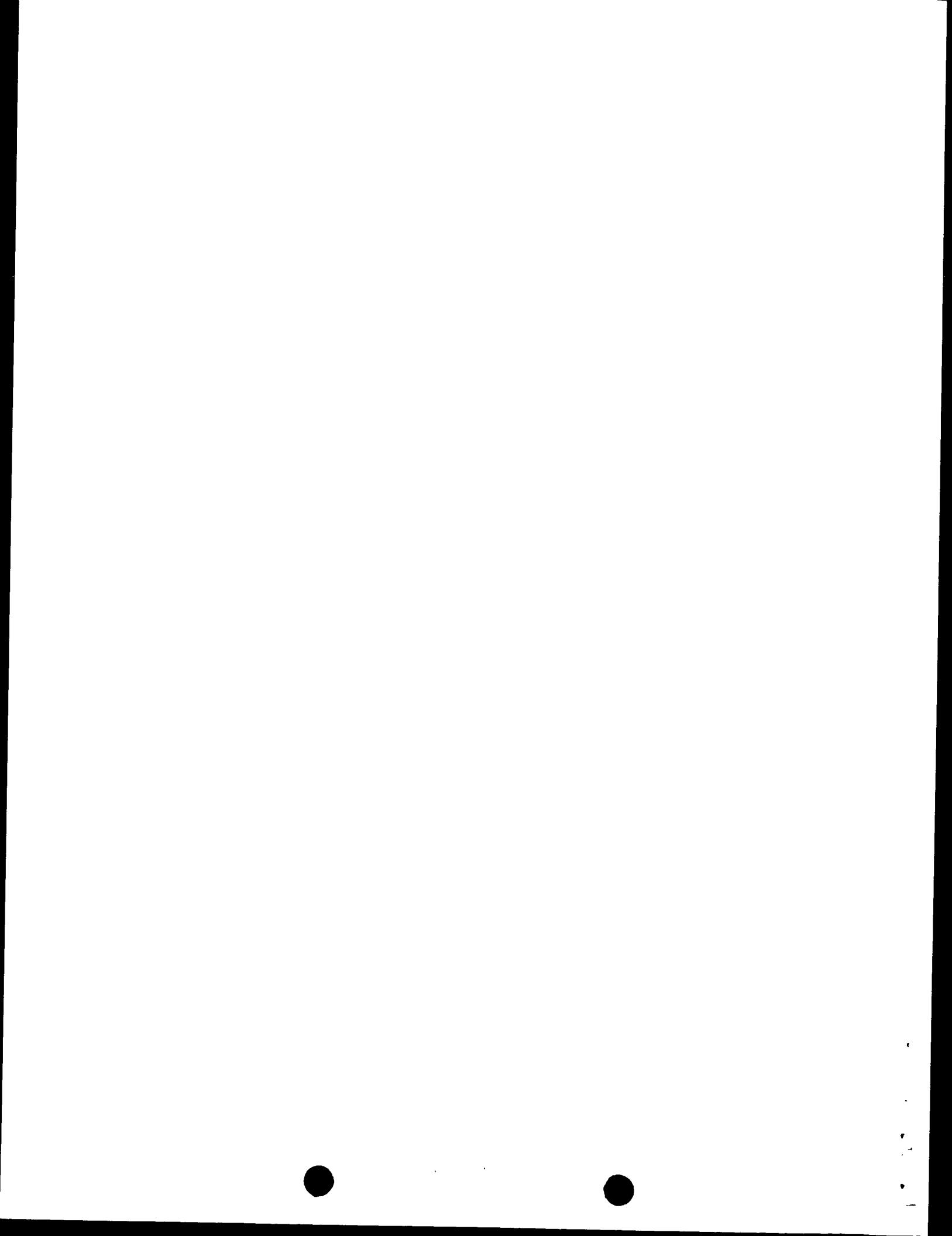
Qy 131 EAKKLMP-E-G-SIVATTYLGKFDEFAVQNYNMGVARASLEANKVYLALDGPDNIRVNAI 188

Db 265 spgyiatdmvkairq-d-vldkivatipvkrlpeiasicawlsseesfstgadfs1 322

Qy 189 SAGPIRTLSAKGFGENTILKEEERAPLKRNYDQEVEGKTAAYLLSDLSSGTGENIHV 248

Db 323 nggih 327

Qy 249 DSGFH 253







Db 158 VKLYHSEAPT 168  
 Qy 87 IDGYVHSIAFA 97

**RESULT 5**  
**ID US-08-121-714-3** STANDARD; PRT; 418 AA.

XX :||| :  
 AC :||| :  
 XX DT 01-JAN-1900  
 XX Sequence 3, Application US/08121714  
 CC GENERAL INFORMATION:  
 CC APPLICANT: Sager, Ruth  
 CC TITLE OF INVENTION: TUMOR SUPPRESSING ACTIVITY  
 CC NUMBER OF SEQUENCES: 8  
 CC CORRESPONDENCE ADDRESS:  
 CC ADDRESSEE: Fish & Richardson  
 CC STREET: 225 Franklin Street  
 CC CITY: Boston  
 CC STATE: Massachusetts  
 CC COUNTRY: U.S.A.  
 CC ZIP: 02110-2804  
 CC COMPUTER READABLE FORM:  
 CC MEDIUM TYPE: 3.5" Diskette, 1.44 Mb  
 CC COMPUTER: IBM PS/2 Model 50Z or 55SX  
 CC OPERATING SYSTEM: MS-DOS (version 5.0)  
 CC SOFTWARE: Wordperfect (Version 5.1)  
 CC CURRENT APPLICATION DATA:  
 CC APPLICATION NUMBER: US/08/121-714  
 CC FILING DATE:  
 CC CLASSIFICATION: 536  
 CC PRIOR APPLICATION DATA:  
 CC APPLICATION NUMBER: 07/938, 823  
 CC FILING DATE: 09/01/92  
 CC APPLICATION NUMBER: 07/844, 296  
 CC FILING DATE: 02/28/92  
 CC APPLICATION NUMBER: 07/662, 216  
 CC FILING DATE: 02/28/91  
 CC ATTORNEY/AGENT INFORMATION:  
 CC NAME: Fraser, Janis K.  
 CC REGISTRATION NUMBER: 34, 819  
 CC REFERENCE/DOCKET NUMBER: 00530/072001  
 CC TELECOMMUNICATION INFORMATION:  
 CC TELEPHONE: (617) 542-5070  
 CC TELEFAX: (617) 542-8906  
 CC TELEX: 200154  
 CC INFORMATION FOR SEQ ID NO: 3:  
 CC SEQUENCE CHARACTERISTICS:  
 CC LENGTH: 418  
 CC TYPE: amino acid  
 CC STRANDEDNESS:  
 CC TOPOLOGY: linear  
 CC SEQUENCE 418 AA: 46736 MW: 945384 CN:

Query Match 5.7%; Score 100; DB 5; Length 418;  
 Best Local Similarity 25.4%; Pred. No. 6.79e+00;  
 Matches 18; Conservative 21; Mismatches 30; Indels 2; Gaps 2;

Db 100 ILEGINFNT-EIPPAQIHGGFQELLRTNQDPSL-QLTGNGFLFSCLKLVDKFLED 157  
 Qy :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||

RESULT 6  
 ID PCT-US95-16930-4 STANDARD; PRT; 764 AA.  
 XX Sequence 4, Application PC/TUS9516930.  
 CC GENERAL INFORMATION:  
 CC APPLICANT: UNIVERSITY OF MASSACHUSETTS MEDICAL SCHOOL  
 CC TITLE OF INVENTION: HETEROLOGOUS POLYPEPTIDE  
 CC TITLE OF INVENTION: PRODUCTION IN THE ABSENCE OF  
 CC NUMBER OF SEQUENCES: 6  
 CC CORRESPONDENCE ADDRESS:  
 CC STREET: 225 Franklin Street Suite 3100  
 CC CITY: Boston  
 CC STATE: MA  
 CC COUNTRY: USA  
 CC ZIP: 02110-2804  
 CC COMPUTER READABLE FORM:  
 CC MEDIUM TYPE: Floppy disk  
 CC COMPUTER: IBM PC compatible  
 CC OPERATING SYSTEM: PC-DOS/MS-DOS  
 CC SOFTWARE: PatentIn Release #.0, Version #.1.30  
 CC CURRENT APPLICATION DATA:  
 CC APPLICATION NUMBER: PCT/US95/16930  
 CC FILING DATE: 27-DEC-1995  
 CC PRIORITY APPLICATION DATA:  
 CC APPLICATION NUMBER: US 08/375, 300  
 CC FILING DATE: 20-JAN-1995  
 CC ATTORNEY/AGENT INFORMATION:  
 CC NAME: Passe, J. Peter  
 CC REGISTRATION NUMBER: 32, 983  
 CC REFERENCE/DOCKET NUMBER: 04020/046W001  
 CC TELECOMMUNICATION INFORMATION:  
 CC LENGTH: 764 amino acids  
 CC TYPE: amino acid  
 CC STRANDEDNESS: single  
 CC TOPOLOGY: linear  
 CC MOLECULE TYPE: protein  
 CC SEQUENCE 764 AA: 8946 MW: 2913896 CN;  
 CC Query Match 5.7%; Score 100; DB 13; Length 764;  
 CC Best Local Similarity 29.1%; Pred. No. 6.79e+00;  
 CC Matches 25; Conservative 25; Mismatches 25; Indels 6; Gaps 6;  
 Db 307 LTKVLLGGIISYRDFVIRCDQVENIGERL-INDIGONMRHSIVRYLITEIFN-FEMI 364  
 Qy 24 VAKVLDQL-GAKLVFTYRKERSKELEKLEQLNQPEAHLYQI-DYQSDEEVINGFEQI 80  
 Db 365 KSDV-LLDITYHIIIRFGHINQNPFP 389  
 Qy 81 GRDVGNIDGVYHSIAFANNEDLGRF 106  
 RESULT 7  
 ID US-08-375-300-4 STANDARD; PRT; 764 AA.  
 XX AC xxxxxx  
 XX DT 01-JAN-1900

DE Sequence 4, Application US/08375300.

XX Sequence 4, Application US/08375300.

CC Patent No. 5679566

GENERAL INFORMATION:

CC APPLICANT: Feng, He

CC ADDRESS: Jacobson, Allan S.

CC TITLE OF INVENTION: HETEROLOGOUS POLYPEPTIDE PRODUCTION IN

CC TITLE OF INVENTION: THE ABSENCE OF NONSENSE-MEDIATED mRNA DECAY FUNCT

CC ION NUMBER OF SEQUENCES: 6

CC CORRESPONDENCE ADDRESS:

CC ADDRESS: Fish &amp; Richardson

CC STREET: 225 Franklin Street Suite 3100

CC CITY: Boston

CC STATE: MA

CC COUNTRY: USA

CC ZIP: 02110-2804

CC COMPUTER READABLE FORM:

CC MEDIUM TYPE: Floppy disk

CC COMPUTER: IBM PC compatible

CC OPERATING SYSTEM: PC-DOS/MS-DOS

CC SOFTWARE: PatentIn Release #1.0, Version #1.30

CC CURRENT APPLICATION DATA:

CC APPLICATION NUMBER: US/08/375,300

CC FILING DATE: 20-JAN-1995

CC ATTORNEY/AGENT INFORMATION:

CC NAME: Fasse, J. Peter

CC REGISTRATION NUMBER: 32,983

CC REFERENCE/DOCKET NUMBER: 04-020/046W01

CC TELECOMMUNICATION INFORMATION:

CC APPLICATION NUMBER: US/08/375,300

CC FILING DATE: 20-JAN-1995

CC ATTORNEY/AGENT INFORMATION:

CC NAME: Fasse, J. P.

CC REGISTRATION NUMBER: 32,983

CC REFERENCE/DOCKET NUMBER: 04-020/046001

CC TELECOMMUNICATION INFORMATION:

CC TELEPHONE: (617)542-5070

CC TELEFAX: (617)542-8906

CC TELEX: 200154

CC MOLECULE TYPE: protein

CC SEQUENCE FOR SEQ ID NO: 2:

CC SEQUENCE CHARACTERISTICS:

CC LENGTH: 1089 amino acids

CC TYPE: amino acid

CC STRANDEDNESS: not relevant

CC TOPOLOGY: linear

CC MOLECULE TYPE: protein

CC SEQUENCE FOR SEQ ID NO: 4:

CC SEQUENCE CHARACTERISTICS:

CC LENGTH: 764 amino acids

CC TYPE: amino acid

CC STRANDEDNESS: single

CC TOPOLOGY: linear

CC MOLECULE TYPE: Protein

CC SEQUENCE FOR SEQ ID NO: 4:

CC SEQUENCE CHARACTERISTICS:

CC LENGTH: 764 amino acids

CC TYPE: amino acid

CC STRANDEDNESS: single

CC TOPOLOGY: linear

CC MOLECULE TYPE: protein

CC SEQUENCE FOR SEQ ID NO: 4:

CC SEQUENCE CHARACTERISTICS:

CC LENGTH: 764 amino acids

CC TYPE: amino acid

CC STRANDEDNESS: single

CC TOPOLOGY: linear

CC MOLECULE TYPE: protein

CC SEQUENCE FOR SEQ ID NO: 4:

CC SEQUENCE CHARACTERISTICS:

CC LENGTH: 764 amino acids

CC TYPE: amino acid

CC STRANDEDNESS: single

CC TOPOLOGY: linear

CC MOLECULE TYPE: protein

CC SEQUENCE FOR SEQ ID NO: 4:

CC SEQUENCE CHARACTERISTICS:

CC LENGTH: 764 amino acids

CC TYPE: amino acid

CC STRANDEDNESS: single

CC TOPOLOGY: linear

CC MOLECULE TYPE: protein

CC SEQUENCE FOR SEQ ID NO: 4:

CC SEQUENCE CHARACTERISTICS:

CC LENGTH: 764 amino acids

CC TYPE: amino acid

CC STRANDEDNESS: single

CC TOPOLOGY: linear

CC MOLECULE TYPE: protein

CC SEQUENCE FOR SEQ ID NO: 4:

CC SEQUENCE CHARACTERISTICS:

CC LENGTH: 764 amino acids

CC TYPE: amino acid

CC STRANDEDNESS: single

CC TOPOLOGY: linear

CC MOLECULE TYPE: protein

CC SEQUENCE FOR SEQ ID NO: 4:

CC SEQUENCE CHARACTERISTICS:

CC LENGTH: 764 amino acids

CC TYPE: amino acid

CC STRANDEDNESS: single

CC TOPOLOGY: linear

CC MOLECULE TYPE: protein

CC SEQUENCE FOR SEQ ID NO: 4:

CC SEQUENCE CHARACTERISTICS:

CC LENGTH: 764 amino acids

CC TYPE: amino acid

CC STRANDEDNESS: single

CC TOPOLOGY: linear

CC MOLECULE TYPE: protein

CC SEQUENCE FOR SEQ ID NO: 4:

CC SEQUENCE CHARACTERISTICS:

CC LENGTH: 764 amino acids

CC TYPE: amino acid

CC STRANDEDNESS: single

CC TOPOLOGY: linear

CC MOLECULE TYPE: protein

CC SEQUENCE FOR SEQ ID NO: 4:

CC SEQUENCE CHARACTERISTICS:

CC LENGTH: 764 amino acids

CC TYPE: amino acid

CC STRANDEDNESS: single

CC TOPOLOGY: linear

CC MOLECULE TYPE: protein

CC SEQUENCE FOR SEQ ID NO: 4:

CC SEQUENCE CHARACTERISTICS:

CC LENGTH: 764 amino acids

CC TYPE: amino acid

CC STRANDEDNESS: single

CC TOPOLOGY: linear

CC MOLECULE TYPE: protein

CC SEQUENCE FOR SEQ ID NO: 4:

CC SEQUENCE CHARACTERISTICS:

CC LENGTH: 764 amino acids

CC TYPE: amino acid

CC STRANDEDNESS: single

CC TOPOLOGY: linear

CC MOLECULE TYPE: protein

CC SEQUENCE FOR SEQ ID NO: 4:

CC SEQUENCE CHARACTERISTICS:

CC LENGTH: 764 amino acids

CC TYPE: amino acid

CC STRANDEDNESS: single

CC TOPOLOGY: linear

CC MOLECULE TYPE: protein

CC SEQUENCE FOR SEQ ID NO: 4:

CC SEQUENCE CHARACTERISTICS:

CC LENGTH: 764 amino acids

CC TYPE: amino acid

CC STRANDEDNESS: single

CC TOPOLOGY: linear

CC MOLECULE TYPE: protein

CC SEQUENCE FOR SEQ ID NO: 4:

CC SEQUENCE CHARACTERISTICS:

CC LENGTH: 764 amino acids

CC TYPE: amino acid

CC STRANDEDNESS: single

CC TOPOLOGY: linear

CC MOLECULE TYPE: protein

CC SEQUENCE FOR SEQ ID NO: 4:

CC SEQUENCE CHARACTERISTICS:

CC LENGTH: 764 amino acids

CC TYPE: amino acid

CC STRANDEDNESS: single

CC TOPOLOGY: linear

CC MOLECULE TYPE: protein

CC SEQUENCE FOR SEQ ID NO: 4:

CC SEQUENCE CHARACTERISTICS:

CC LENGTH: 764 amino acids

CC TYPE: amino acid

CC STRANDEDNESS: single

CC TOPOLOGY: linear

CC MOLECULE TYPE: protein

CC SEQUENCE FOR SEQ ID NO: 4:

CC SEQUENCE CHARACTERISTICS:

CC LENGTH: 764 amino acids

CC TYPE: amino acid

CC STRANDEDNESS: single

CC TOPOLOGY: linear

CC MOLECULE TYPE: protein

CC SEQUENCE FOR SEQ ID NO: 4:

CC SEQUENCE CHARACTERISTICS:

CC LENGTH: 764 amino acids

CC TYPE: amino acid

CC STRANDEDNESS: single

CC TOPOLOGY: linear

CC MOLECULE TYPE: protein

CC SEQUENCE FOR SEQ ID NO: 4:

CC SEQUENCE CHARACTERISTICS:

CC LENGTH: 764 amino acids

CC TYPE: amino acid

CC STRANDEDNESS: single

CC TOPOLOGY: linear

CC MOLECULE TYPE: protein

CC SEQUENCE FOR SEQ ID NO: 4:

CC SEQUENCE CHARACTERISTICS:

CC LENGTH: 764 amino acids

CC TYPE: amino acid

CC STRANDEDNESS: single

CC TOPOLOGY: linear

CC MOLECULE TYPE: protein

CC SEQUENCE FOR SEQ ID NO: 4:

CC SEQUENCE CHARACTERISTICS:

CC LENGTH: 764 amino acids

CC TYPE: amino acid

CC STRANDEDNESS: single

CC TOPOLOGY: linear

CC MOLECULE TYPE: protein

CC SEQUENCE FOR SEQ ID NO: 4:

CC SEQUENCE CHARACTERISTICS:

CC LENGTH: 764 amino acids

CC TYPE: amino acid

CC STRANDEDNESS: single

CC TOPOLOGY: linear

CC MOLECULE TYPE: protein

CC SEQUENCE FOR SEQ ID NO: 4:

CC SEQUENCE CHARACTERISTICS:

CC LENGTH: 764 amino acids

CC TYPE: amino acid

CC STRANDEDNESS: single

CC TOPOLOGY: linear

CC MOLECULE TYPE: protein

CC SEQUENCE FOR SEQ ID NO: 4:

CC SEQUENCE CHARACTERISTICS:

CC LENGTH: 764 amino acids

CC TYPE: amino acid

CC STRANDEDNESS: single

CC TOPOLOGY: linear

CC MOLECULE TYPE: protein

CC SEQUENCE FOR SEQ ID NO: 4:

CC SEQUENCE CHARACTERISTICS:

CC LENGTH: 764 amino acids

CC TYPE: amino acid

CC STRANDEDNESS: single

CC TOPOLOGY: linear

CC MOLECULE TYPE: protein

CC SEQUENCE FOR SEQ ID NO: 4:

CC SEQUENCE CHARACTERISTICS:

CC LENGTH: 764 amino acids

CC TYPE: amino acid

CC STRANDEDNESS: single

CC TOPOLOGY: linear

CC MOLECULE TYPE: protein

CC SEQUENCE FOR SEQ ID NO: 4:

CC SEQUENCE CHARACTERISTICS:

CC LENGTH: 764 amino acids

CC TYPE: amino acid

CC STRANDEDNESS: single

CC TOPOLOGY: linear

CC MOLECULE TYPE: protein

CC SEQUENCE FOR SEQ ID NO: 4:

CC SEQUENCE CHARACTERISTICS:

CC LENGTH: 764 amino acids

CC TYPE: amino acid

CC STRANDEDNESS: single

CC TOPOLOGY: linear

CC MOLECULE TYPE: protein

CC SEQUENCE FOR SEQ ID NO: 4:

CC SEQUENCE CHARACTERISTICS:

CC LENGTH: 764 amino acids

CC TYPE: amino acid

CC STRANDEDNESS: single

CC TOPOLOGY: linear

CC MOLECULE TYPE: protein

CC SEQUENCE FOR SEQ ID NO: 4:

CC SEQUENCE CHARACTERISTICS:

CC LENGTH: 764 amino acids

CC TYPE: amino acid

CC STRANDEDNESS: single

CC TOPOLOGY: linear

CC MOLECULE TYPE: protein

CC SEQUENCE FOR SEQ ID NO: 4:

CC SEQUENCE CHARACTERISTICS:

CC LENGTH: 764 amino acids

CC TYPE: amino acid

CC STRANDEDNESS: single

CC TOPOLOGY: linear

CC MOLECULE TYPE: protein

CC SEQUENCE FOR SEQ ID NO: 4:

CC SEQUENCE CHARACTERISTICS:

CC LENGTH: 764 amino acids

CC TYPE: amino acid

CC STRANDEDNESS: single

CC TOPOLOGY: linear

CC MOLECULE TYPE: protein

CC SEQUENCE FOR SEQ ID NO: 4:

CC SEQUENCE CHARACTERISTICS:

CC LENGTH: 764 amino acids

CC TYPE: amino acid

CC STRANDEDNESS: single

CC TOPOLOGY: linear

CC MOLECULE TYPE: protein

CC SEQUENCE FOR SEQ ID NO: 4:

CC SEQUENCE CHARACTERISTICS:

CC LENGTH: 764 amino acids

COMPUTER READABLE FORM:  
 CC MEDIUM TYPE: FLOPPY disk  
 CC COMPUTER: IBM PC compatible  
 CC OPERATING SYSTEM: PC-DOS/MS-DOS  
 CC CURRENT APPLICATION DATA:  
 CC SOFTWARE: PatentIn Release #1.0, Version #1.30  
 CC APPLICATION NUMBER: US/08/375, 300  
 CC FILING DATE: 20-JAN-1995  
 CC ATTORNEY/AGENT INFORMATION:  
 CC NAME: Fasse, J. P.  
 CC REGISTRATION NUMBER: 32,983  
 CC REFERENCE/DOCKET NUMBER: 04020/046001  
 CC TELECOMMUNICATION INFORMATION:  
 CC TELEPHONE: (617)542-2070  
 CC TELEFAX: (617)542-9006  
 CC TELLEX: 200154  
 CC INFORMATION FOR SEQ ID NO: 2:  
 CC SEQUENCE CHARACTERISTICS:  
 CC LENGTH: 1089 amino acids  
 CC TYPE: amino acid  
 CC STRANDEDNESS: not relevant  
 CC TOPOLOGY: linear  
 CC MOLECULE TYPE: protein  
 CC SEQUENCE 1089 AA: 126746 MW: 5973553 CN;

Query Match 5.7% Score 100; DB 7; Length 1089;  
 Best Local Similarity 29.1%; Pred. No. 6.79e+00; Gaps 6;  
 Matches 25; Conservative 25; MisMatches 30; Indels 6; Gaps 6;

Db 632 LTKVLLGGISYRDRFVICDQLENLERGLIE INDYQNNHRISNVRYLTIEFN-FEMI 689  
 QY 24 VAKVLQQL-GAKLVEFTKRSRKELQLEKLQLQNQPEAHLYQQI-DVQSDEEVINGFEQI 80

Db 690 KSDV-LDDTIYHIFRGINNQNPF 714  
 QY 81 GRDVGNTIDGVYHSIAFANMEDLRGRF 106

RESULT 10  
 ID US-08-002-202-9 STANDARD: PRT; 394 AA.  
 XX XXXXXXXX  
 XX DT 01-JAN-1900  
 XX DE Sequence 9, Application US/08002202.

Sequence 9, Application US/08002202.  
 Patent No. 5604201  
 GENERAL INFORMATION:  
 CC APPLICANT: Thomas, Laurel  
 CC APPLICANT: Anderson, Eric D  
 CC APPLICANT: Thomas, Garry  
 CC APPLICANT: Hayflck, Joel S  
 CC TITLE OF INVENTION: Methods and Reagents for Inhibiting  
 CC TITLE OF INVENTION: Furin Endoprotease  
 CC NUMBER OF SEQUENCES: 21  
 CC CORRESPONDENCE ADDRESS:  
 CC ADDRESSEE: Allegretti and Witcoff, Ltd.  
 CC STREET: 10 South Wacker Drive, Suite 3000  
 CC CITY: Chicago  
 CC STATE: Illinois  
 CC COUNTRY: USA  
 CC ZIP: 60606

CC COMPUTER READABLE FORM:  
 CC MEDIUM TYPE: Floppy disk  
 CC COMPUTER: IBM PC compatible  
 CC OPERATING SYSTEM: PC-DOS/MS-DOS  
 CC SOFTWARE: PatentIn Release #1.0, Version #1.25  
 CC CURRENT APPLICATION DATA:  
 CC APPLICATION NUMBER: US/08/002, 202  
 CC FILING DATE: 08-JAN-1993  
 CC CLASSIFICATION: 514  
 CC ATTORNEY/AGENT INFORMATION:

ATTORNEY/AGENT INFORMATION:  
 CC NAME: No. 5604201ian, Kevin E  
 CC REGISTRATION NUMBER: 35,30003  
 CC REFERENCE/DOCKET NUMBER: 92,448  
 CC TELECOMMUNICATION INFORMATION:  
 CC TELEPHONE: 312-715-1000  
 CC TELEX: 312-715-1234  
 CC TELEX: 910-221-5317  
 CC INFORMATION FOR SEQ ID NO: 9:  
 CC SEQUENCE CHARACTERISTICS:  
 CC LENGTH: 394 amino acids  
 CC TYPE: amino acid  
 CC TOPOLOGY: linear  
 CC MOLECULE TYPE: protein  
 CC FEATURE:  
 CC NAME/KEY: Protein  
 CC LOCATION: 1-394  
 CC OTHER INFORMATION: /label= Variant  
 CC OTHER INFORMATION: /note= "This amino acid sequence is the amino acid sequence of the modified alpha-1-antitrypsin protein, alpha-1-antitrypsin Portland;"  
 CC id  
 CC OTHER INFORMATION: sequence of the modified alpha-1-antitrypsin protein, alpha-1-antitrypsin Portland;  
 CC OTHER INFORMATION: protein, alpha-1-antitrypsin  
 CC SQ SEQUENCE 394 AA: 44432 MW: 838751 CN;  
 Query Match 5.6% Score 99; DB 7; Length 394;  
 Best Local Similarity 25.4%; Pred. No. 7.82e+00;  
 Matches 20; Conservative 20; MisMatches 31; Indels 2; Gaps 2;  
 Db 76 :ILEGILNFIT-EIEPAQIHEGFOELLRTLNOPDSQL-QLTGNGFLSQLGLKVDFLED 133  
 QY 27 VLQLGAKLVFTYKERSKELKELELQINOPEARHYQDVSDEEVINGFEQIKGDYN 86  
 Db 134 VKKLYHSEFT 144  
 QY 87 :V:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||

CC COMPUTER READABLE FORM:  
 CC MEDIUM TYPE: Floppy disk  
 CC COMPUTER: IBM PC compatible  
 CC OPERATING SYSTEM: PC-DOS/MS-DOS  
 CC SOFTWARE: PatentIn Release #1.0, Version #1.25  
 CC CURRENT APPLICATION DATA:  
 CC APPLICATION NUMBER: US/08/002, 202  
 CC FILING DATE: 08-JAN-1993  
 CC CLASSIFICATION: 514  
 CC ATTORNEY/AGENT INFORMATION:

CC COMPUTER READABLE FORM:  
 CC MEDIUM TYPE: Floppy disk  
 CC COMPUTER: IBM PC compatible  
 CC OPERATING SYSTEM: PC-DOS/MS-DOS  
 CC SOFTWARE: PatentIn Release #1.0, Version #1.25  
 CC CURRENT APPLICATION DATA:  
 CC APPLICATION NUMBER: US/08/002, 202  
 CC FILING DATE: 08-JAN-1993  
 CC CLASSIFICATION: 514  
 CC ATTORNEY/AGENT INFORMATION:

NAME: No. 5604201nan, Kevin E  
 REGISTRATION NUMBER: 35,30003  
 REFERENCE/DOCKET NUMBER: 92,448  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 312-715-1100  
 TELEX: 3112-715-1234  
 TELFAX: 910-221-5317  
 INFORMATION FOR SEQ ID NO: 11:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 394 amino acids  
 TYPE: amino acid  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 FEATURE:  
 NAME/KEY: Protein  
 LOCATION: 1..394  
 OTHER INFORMATION: /label= Variants  
 OTHER INFORMATION: /note= "This amino acid sequence is the amino ac  
 id  
 OTHER INFORMATION: sequence of the modified alpha-1-antitrypsin  
 OTHER INFORMATION: variant, alpha-1-antitrypsin Pittsburgh"  
 SEQUENCE 394 AA; 44347 MW; 838436 CN;  
 Query Match 5.6%; Score 99; DB 7; Length 394;  
 Best Local Similarity 25.4%; Pred. No. 7.82e+00;  
 Matches 18; Conservative 20; Mismatches 31; Indels 2; Gaps 2;  
 RESULT 13  
 Db 76 ILEGIFNFI-ETPEAQIHGFOELLRLTINQPSQLKLVDKFLED 133  
 QY 27 VLDQGAKLVFTYRKERSKELLEQLNQPEAHLYQIDYQSDEEVINGFEQIGKDVGN 86  
 Db ~ 134 VKLHSEAF 144  
 QY 87 IDGVHSIAFA 97  
 RESULT 13  
 Db 76 ILEGIFNFI-ETPEAQIHGFOELLRLTINQPSQLKLVDKFLED 133  
 QY 27 VLDQGAKLVFTYRKERSKELLEQLNQPEAHLYQIDYQSDEEVINGFEQIGKDVGN 86  
 Db ~ 134 VKLHSEAF 144  
 QY 87 IDGVHSIAFA 97  
 Sequence 6, Application US/08002202.  
 Sequence 6, Application US/08002202-6  
 XX ID US-08-002-202-6 STANDARD; PRT; 394 AA.  
 XX AC XXXXXX  
 DT 01-JAN-1900  
 XX  
 CORRESPONDENCE ADDRESS:  
 CC Pat. No. 5604201  
 CC GENERAL INFORMATION:  
 CC APPLICANT: Haylick, Joel S  
 CC APPLICANT: Thomas, Garry  
 CC APPLICANT: Anderson, Eric D  
 CC APPLICANT: Thomas, Laurel  
 CC TITLE OF INVENTION: Methods and Reagents for Inhibiting  
 CC TITLE OF INVENTION: Furin Endoprotease  
 CC NUMBER OF SEQUENCES: 21  
 CC CORRESPONDENCE ADDRESS:  
 CC STREET: 10 South Wacker Drive, Suite 3000  
 CC CITY: Chicago  
 CC STATE: Illinois  
 CC ZIP: 60606  
 CC COMPUTER READABLE FORM:  
 CC MEDIUM TYPE: Floppy disk  
 CC COMPUTER: IBM PC compatible  
 CC OPERATING SYSTEM: PC-DOS/MS-DOS  
 CC SOFTWARE: PatentIn Release #1.0, Version #1.25  
 CC CURRENT APPLICATION DATA:  
 CC APPLICATION NUMBER: US/08/002,202  
 CC FILING DATE: 08-JAN-1993  
 CC CLASSIFICATION: 514  
 CC ATTORNEY/AGENT INFORMATION:  
 CC NAME: No. 5604201nan, Kevin E  
 CC LENGTH: 647 amino acids

CC TYPE: amino acid  
 CC STRANDEDNESS: single  
 CC TOPOLOGY: linear  
 CC MOLECULE TYPE: Protein  
 SQ SEQUENCE: 647 AA; 72413 MW; 2325170 CN;  
 Query Match 5.5%; Score 97; DB 7; Length 647;  
 Best Local Similarity 28.9%; Pred. No. 1.04e+01;  
 Matches 26; Conservative 23; Mismatches 34; Indels 7; Gaps 6;  
 Db 249 VGNNSAPIVPTLDDFRSMAFTGIFRSRPHEDHDLAGEETASV-TYGPANAK-VPHSA 305  
 Qy 84 VGNIDGVY-HSIA-FAANMEDLRGRSETSRSEGF-DLAQDISSYSLTIVAEAKKLMPPEG 140  
 Db 306 S-SDTSLISAYSGPSYSSLSITTSSTEAK 34  
 Qy 141 SIVATYLGSEFFAVONYNTNGVAKASLEAN 170  
 [REDACTED] 14  
 US-08-292-945-6 STANDARD; PRT; 201 AA.  
 AC xxxxxxxx  
 XX DT 01-JAN-1900  
 XX DE Sequence 6, Application US/08292945.  
 XX Sequence 6, Application US/08292945  
 CC Patent No. 5585478  
 CC GENERAL INFORMATION:  
 CC APPLICANT: Lim, Bing  
 CC ADDRESS: Kilpatrick & Cody  
 CC STREET: 1100 Peachtree Street, Suite 2800  
 CC CITY: Atlanta  
 CC STATE: Georgia  
 CC COUNTRY: United States  
 CC ZIP: 30303-4530  
 COMPUTER READABLE FORM  
 CC MEDIUM TYPE: Floppy disk  
 CC COMPUTER: IBM PC compatible  
 CC OPERATING SYSTEM: PC-DOS/MS-DOS  
 CC SOFTWARE: Patentin Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 CC APPLICATION NUMBER: US/08/292 945  
 CC FILING DATE:  
 CC CLASSIFICATION: 435  
 CC PRIOR APPLICATION DATA:  
 CC APPLICATION NUMBER: US 07/990 337  
 CC FILING DATE: December 10, 1992  
 CC ATTORNEY/AGENT INFORMATION:  
 CC NAME: Pabst, Patrea L.  
 CC REGISTRATION NUMBER: 31,284  
 CC TELEPHONE: (404) 815-6508  
 CC TELEFAX: (404) 815-6555  
 CC INFORMATION FOR SEQ ID NO: 6:  
 CC SEQUENCE CHARACTERISTICS:  
 CC LENGTH: 201 amino acids  
 CC STRANDEDNESS: single  
 CC TOPOLOGY: linear  
 CC MOLECULE TYPE: protein  
 CC HYPOTHETICAL: NO  
 CC ANTI-SENSE: NO  
 CC FRAGMENT TYPE: N-terminal  
 CC ORIGINAL SOURCE:  
 CC ORGANISM: Murine  
 SQ SEQUENCE 201 AA; 22919 MW; 225152 CN;  
 Query Match 5.2%; Score 91; DB 11; Length 201;  
 Best Local Similarity 23.0%; Pred. No. 2.40e+01;  
 Matches 28; Conservative 27; Mismatches 63; Indels 4; Gaps 4;  
 Db 15 DDLSKLNKPQQPKSLKELOEM-DKDDESLTK-YKTKLGDYPVVADTPNVTVRILS 72  
 Qy 29 DQLGAKLVFTYRKERSRKELQLEQNLQPEAHLYQIDVSDEEVINGEQIGKDVGND 88  
 Db 73 LYCDSAPPITMDLTGDLALKTDFVLEKIE-YRKVINFVKNDIVS-GLKYQHTYR 130  
 Qy 89 GYVSIATFAFNMEDLRGRFSETSRSEGF-LAQDISSYSLTIVAEAKKLMPEGGSIVATTYL 148  
 Db 131 TG 132  
 Qy 149 GG 150

RESULT 15  
 ID PCT-US93-12074-6 STANDARD;  
 PRT; 201 AA.

XX DT 01-JAN-1900  
 XX DE Sequence 6, Application PC/TUSS9312074.  
 XX Sequence 6, Application PC/TUSS9312074  
 CC GENERAL INFORMATION:  
 CC APPLICANT:  
 CC TITLE OF INVENTION: D4 Gene and Methods of Use Thereof  
 CC NUMBER OF SEQUENCES: 8  
 CC COMPUTER READABLE FORM:  
 CC MEDIUM TYPE: Floppy disk  
 CC COMPUTER: IBM PC compatible  
 CC OPERATING SYSTEM: PC-DOS/MS-DOS  
 CC SOFTWARE: Patentin Release #1.0, Version #1.25 (EPO)  
 CURRENT APPLICATION DATA:  
 CC APPLICATION NUMBER: PCT/US93/12074  
 CC FILING DATE:  
 CC PRIORITY APPLICATION DATA:  
 CC APPLICATION NUMBER: US 07/990,337  
 CC FILING DATE: 10-DEC-1992  
 CC INFORMATION FOR SEQ ID NO: 6:  
 CC SEQUENCE CHARACTERISTICS:  
 CC LENGTH: 201 amino acids  
 CC TYPE: amino acid  
 CC STRANDEDNESS: single  
 CC TOPOLOGY: Linear  
 CC MOLECULE TYPE: Protein  
 CC HYPOTHETICAL: NO  
 CC ANTI-SENSE: NO  
 CC FRAGMENT TYPE: N-terminal  
 CC ORIGINAL SOURCE:  
 CC ORGANISM: Murine  
 SQ SEQUENCE 201 AA; 22919 MW; 225152 CN;  
 Query Match 5.2%; Score 91; DB 11; Length 201;  
 Best Local Similarity 23.0%; Pred. No. 2.40e+01;  
 Matches 28; Conservative 27; Mismatches 63; Indels 4; Gaps 4;  
 Db 15 DDLSKLNKPQQPKSLKELOEM-DKDDESLTK-YKTKLGDYPVVADTPNVTVRILS 72  
 Qy 29 DQLGAKLVFTYRKERSRKELQLEQNLQPEAHLYQIDVSDEEVINGEQIGKDVGND 88  
 Db 73 LYCDSAPPITMDLTGDLALKTDFVLEKIE-YRKVINFVKNDIVS-GLKYQHTYR 130  
 Qy 89 GYVSIATFAFNMEDLRGRFSETSRSEGF-LAQDISSYSLTIVAEAKKLMPEGGSIVATTYL 148  
 Db 131 TG 132

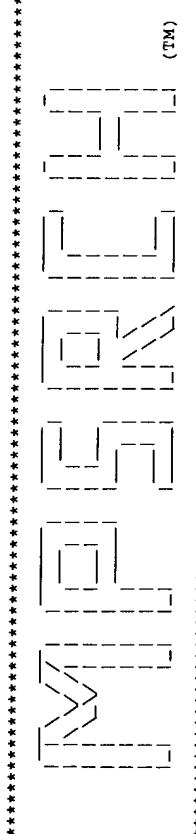
Fri Dec 5 12:31:28 1997

US-08-790-043A-1.rai

Page 8

Qy 149 | GG 150

Search completed: Thu Dec 4 16:11:18 1997  
Job time : 48 secs.

\*\*\*\*\*  
  
\*\*\*\*\*

Release 2.1D John F. Collins Biocomputing Research Unit.  
Copyright (c) 1993, 1994, 1995 University of Edinburgh, U.K.

.

Distribution rights by IntelligentGenetics, Inc.

Mpsrch\_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Thu Dec 4 16:01:38 1997; MasPar time 12.50 Seconds

591.864 Million cell updates/sec  
Boular output not generated.

Title: >US-08-790-043A-1  
(1-256) from US08790043A.pep

Perfect Score: 1753

Sequence: 1 MNLNEKTYVIMGIANRKSTI.....LSSGTGENIHVDGFAIK 256

Scoring table: PAM 150

Gap 11

Searched: 91006 seqs, 28888923 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: pir51

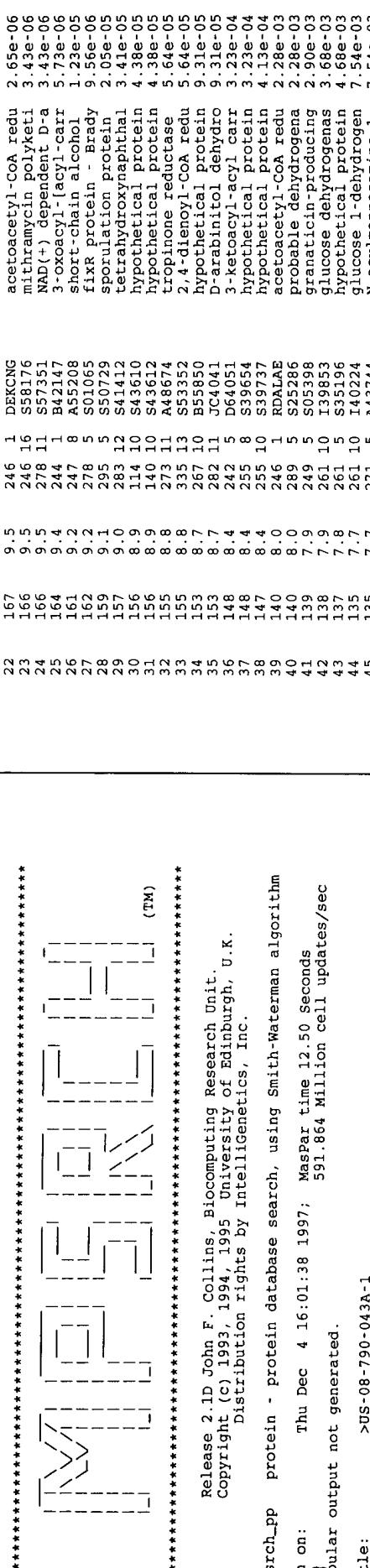
1:ann1 2:ann2 3:ann3 4:ann4 5:unann1 6:unann2 7:unann3  
8:unann4 9:unann5 10:unann6 11:unann7 12:unann8  
13:unann9 14:unann10 15:unenc 16:unrev

Statistics: Mean 47.222; Variance 135.383; scale 0.349

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Rank	No.	Score	Query	Match	Length	DB ID	Description	Pred. No.	Description	Pred.
1	748	42.7		262	9	S48029	short-chain alcohol	2	41e-84	
2	732	41.8		262	9	B43729	envM protein - Salmo	4	5.58e-82	
3	718	41.0		295	9	B64139	short chain alcohol	4	4.49e-80	
4	365	20.8		385	16	S17761	enoyl-acyl carrier p	4	0.05e-31	
5	229	13.1		249	5	B37762	bile acid 7-dehydrox	1	3.0e-13	
6	224	12.8		244	5	JN0703	carboxyl reductase (	5	5.31e-13	
7	225	12.8		249	5	A31841	27K-2 protein (choli	2	0.1e-13	
8	219	12.5		244	5	A28053	adipocyte p27 protein	2	1.5e-12	#label BER
9	214	12.2		260	11	B48674	tropinone reductase	8	6.66e-12	#cross-references NCBI:N-121825;
10	212	12.1		249	5	A28212	27K bile acid 7-dehy	1	5.1e-11	#note sequence extracted from NCBI backbone
11	197	11.2		245	1	C24706	nodulation protein n	9	2.5e-10	SUMMARY #length 262 #molecular-weight 27864 #checksum 2286
12	190	10.8		266	5	A42468	7alpha-hydroxytro	4	1.4e-09	Query Match Score 748; Best Local Similarity 42.7%; Pred. No. 2.41e-84;
13	183	10.4		254	9	S56492	hypothetical protein	3	9.0e-08	Matches 111; Conservative 44.4%; Mismatches 71; Indels 3; Gaps 3;
14	182	10.4		254	9	S34724	probable oxidoreduct	5	2.0e-08	
15	181	10.3		262	16	A48950	afaxton biosynthes	6	7.0e-08	
16	179	10.2		255	5	A38527	7alpha-hydroxysteroi	6	1.5e-07	
17	171	9.8		256	8	A57149	glutonate 5-dehydrot	9	3.0e-07	
18	171	9.8		320	5	S22450	3'-oxoacyl-l-acyl-carr	4	3.8e-07	
19	169	9.6		242	11	S39508	hypothetical protein	1	5.8e-06	
20	169	9.6		261	9	S56475	hypothetical protein	1	5.0e-06	
21	168	9.6		268	11	C48674	tropinone reductase	2	0.04e-06	

\*\*\*\*\*  
  
\*\*\*\*\*

#### ALIGNMENTS

RESULT	1	S48029 #type complete
ENTRY		short-chain alcohol dehydrogenase homolog envM - Escherichia coli
TITLE		enoyl-ACP reductase
ALTERNATE_NAMES		#formal_name Escherichia coli
ORGANISM		DATE
		07-May-1995 #sequence_revision 21-Jul-1995 #text_change
ACCESSIONS		27-Oct-1995
REFERENCE	S48029; A47681	A.R.
REFID		Plant Mol. Biol. (1994) 25:771-790
ORGANISM		The use of a hybrid genetic system to study the functional relationship between prokaryotic and plant multi-enzyme fatty acid synthetase complexes.
DATE		
ACCESSTION	S48029	#accession
RESIDUES		#status preliminary
MOLTYPE		#molecule_type DNA
DNATYPE		#residues 1-162 #label KAT
CROSSREFS		#cross-references EMBL:X78733
REFERENCE	A47681	REFERENCE
REFID		Bergler, H.; Hogenauer, G.; Turnowsky, F.
ORGANISM		J. Gen. Microbiol. (1992) 138:2993-2100
DATE		Sequences of the envM gene and of two mutated alleles in Escherichia coli.
ACCESSTION	A47681	#accession MUID:93123967
MOLTYPE		#status preliminary
DNATYPE		#molecule_type DNA
RESIDUES		#residues 1-162 #label BER
CROSSREFS		#cross-references NCBI:N-121825;
NOTE		#note sequence extracted from NCBI backbone
SUMMARY		#length 262 #molecular-weight 27864 #checksum 2286
MATCH		Query Match Score 748; Best Local Similarity 42.7%; Pred. No. 2.41e-84;
LOCALSIMILARITY		Matches 111; Conservative 44.4%; Mismatches 71; Indels 3; Gaps 3;
PREDNO		
INDELS		
GAPS		
LENTHIT		
ERRHIT		
ERRPCT		
ERRRATIO		
ERRRATIOHIT		
ERRRATIOERR		
ERRRATIOERRHIT		
ERRRATIOERRRATIO		
ERRRATIOERRRATIOHIT		
ERRRATIOERRRATIOERR		
ERRRATIOERRRATIOERRHIT		
ERRRATIOERRRATIOERRRATIO		
ERRRATIOERRRATIOERRRATIOHIT		
ERRRATIOERRRATIOERRRATIOERR		
ERRRATIOERRRATIOERRRATIOERRHIT		
ERRRATIOERRRATIOERRRATIOERRRATIO		
ERRRATIOERRRATIOERRRATIOERRRATIOHIT		
ERRRATIOERRRATIOERRRATIOERRRATIOERR		
ERRRATIOERRRATIOERRRATIOERRRATIOERRHIT		
ERRRATIOERRRATIOERRRATIOERRRATIOERRRATIO		
ERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOHIT		
ERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERR		
ERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRHIT		
ERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIO		
ERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOHIT		
ERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERR		
ERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRHIT		
ERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIO		
ERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOHIT		
ERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERR		
ERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRHIT		
ERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIO		
ERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOHIT		
ERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERR		
ERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRHIT		
ERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIO		
ERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOHIT		
ERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERR		
ERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRHIT		
ERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIO		
ERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOHIT		
ERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERR		
ERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRHIT		
ERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIO		
ERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOHIT		
ERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERR		
ERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRHIT		
ERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIO		
ERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOHIT		
ERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERR		
ERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRHIT		
ERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIO		
ERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOHIT		
ERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERR		
ERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRHIT		
ERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIO		
ERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOHIT		
ERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERR		
ERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRHIT		
ERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIO		
ERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOHIT		
ERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERR		
ERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRHIT		
ERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIO		
ERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOHIT		
ERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERR		
ERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRHIT		
ERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIO		
ERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOHIT		
ERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERR		
ERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRHIT		
ERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIO		
ERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOHIT		
ERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERR		
ERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRHIT		
ERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIO		
ERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOHIT		
ERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERR		
ERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRHIT		
ERRRATIO		
ERRRATIOHIT		
ERRRATIOERR		
ERRRATIOERRHIT		
ERRRATIO		
ERRRATIOHIT		
ERRRATIOERR		
ERRRATIOERRHIT		
ERRRATIO		
ERRRATIOHIT		
ERRRATIOERR		
ERRRATIOERRHIT		
ERRRATIO		
ERRRATIOHIT		
ERRRATIOERR		
ERRRATIOERRHIT		
ERRRATIO		
ERRRATIOHIT		
ERRRATIOERR		
ERRRATIOERRHIT		
ERRRATIO		
ERRRATIOHIT		
ERRRATIOERR		
ERRRATIOERRHIT		
ERRRATIO		
ERRRATIOHIT		
ERRRATIOERR		
ERRRATIOERRHIT		
ERRRATIO		
ERRRATIOHIT		
ERRRATIOERR		
ERRRATIOERRRATIOERRRATIOERRRATIOERRRATIOERRR		

QY 64 QIDVQSDEEVINGFEQIKDVGNNIDGVYHSTAFFANMEDLRGRFSET-SREGFLAQDIS 122  
 DB 122 ysfamakacrtmngpsalitlsygaeraipnvnmgakasleanvrmnanangpeg 181  
 QY 123 YSLTIVAAHEAKRMLPMEGSIVATTYLGEFAVNQNNVMGVAKASLEANVKYLAQDGPDN 182  
 DB 182 vrnaisagpirtlasaagikdrkmahceavtpirrtvtiedvngsaaficsdisagi 241  
 QY 183 IRVNAISAGPIRTLSAKGVGGNTILKEIEERAPLKRNVQDVVEGKTAAYVLSLSSGV 242  
 Db 242 gevvhvddggf 251  
 QY 243 GENIHDVDSGF 252

RESULT 2 B43729 #type complete  
 ENTRY envM protein - Salmonella typhimurium  
 ORGANISM #formal\_name Salmonella typhimurium  
 REFERENCE 03-Mar-1993 #sequence\_revision 03-Mar-1993 #text\_change  
 23-Jun-1993  
 SEQUENCES B43729  
 REFERENCE A43729  
 AUTHORS Turunowsky, F.; Fuchs, K.; Jeschek, C.; Hoegenauer, G.  
 J. Bacteriol. (1989) 171:6555-6565  
 JOURNAL envM genes of Salmonella typhimurium and Escherichia coli.  
 ACCESSION B43729  
 #status Preliminary  
 #molecule\_type DNA  
 #residues 1-262 #label TUR  
 #cross-references GB:M31806  
 SUMMARY #length 262 #molecular\_weight 27761 #checksum 2093  
 Query Match 41.0%; Score 732; DB 9; Length 262;  
 Best Local Similarity 43.8%; Pred. No. 4.58e-82;  
 Matches 110; Conservative 65; Mismatches 71; Indels 5; Gaps 5;  
 Db 4 lsgkrllvgtavksltsiyqamhregalaftqndk1kgvreaafqglssiv-1-61  
 QY 4 LENKTYTIMGANKRSSTAFGVAKVLQDGAKLYFTYKERSKELEKLLQINOPPAHY 63  
 Db 62 pcdavaedisdamfaelg-nwpxpkfdgfvhsigafqdgdyvnnavtregfkvahdis 120  
 QY 64 QIDVQSDEEVINGFEQIKDVGNNIDGVYHSTAFFANMEDLRGRFSET-SREGFLAQDIS 121  
 Db 121 sysfvamakacrtmngpsalitlsygaeraipnvnmgakasleanvrmnanangpe 180  
 QY 122 YSLTIVAAHEAKRMLPMEGSIVATTYLGEFAVNQNNVMGVAKASLEANVKYLAQDGPD 181  
 Db 181 gvrvnaisagpirtlasaagikdrkmahceavtpirrtvtiedvngsaaficsdisagi 240  
 QY 182 IRVNAISAGPIRTLSAKGVGGNTILKEIEERAPLKRNVQDVVEGKTAAYVLSLSSGV 241  
 Db 241 gevvhvddggf 251  
 QY 242 GENIHDVDSGF 252

RESULT 3 B64139 #type complete  
 ENTRY short chain alchohol dehydrogenase homolog (envM) homolog -  
 TITLE Haemophilus influenzae (strain Rd KW20)  
 ORGANISM #formal\_name Haemophilus influenzae  
 DATE 18-Aug-1995 #sequence\_revision 18-Aug-1995 #text\_change  
 10-May-1996  
 ACCESSIONS B64139  
 REFERENCE A64000  
 #authors  
 Flieischmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirchner, E.F.; Keravage, A.R.; Bult, C.J.; Tomb, J.F.; Doughtery, B.A.; Merrick, J.M.; McKenney, K.; Sutton, G.; Fleischmann, R.D.; Fields, C.; Gocayne, J.D.; Scott, J.; Shriray, R.; Liu, L.I.; Glodek, A.; Kelley, J.M.; Weidman, J.F.; Phillips, C.A.; Spriggs, T.; Heblom, E.; Cotton, A.R.

RESULT 4 S11761 #type complete  
 ENTRY enoyl-acyl carrier protein reductase - rape  
 TITLE  
 ORGANISM #formal\_name Brassica napus #common\_name rape  
 DATE 22-Nov-1993 #sequence\_revision 22-Nov-1993 #text\_change  
 22-Nov-1993  
 ACCESSIONS S11761  
 REFERENCE S11761  
 #authors Kater, M.M.; Koningsstein, G.M.; Nijkamp, H.J.J.; Stuitje, A.R.  
 #status Preliminary  
 #molecule\_type KAT  
 #residues 1-385 #label KAT  
 SUMMARY #length 385 #molecular\_weight 40486 #checksum 4099  
 Query Match 20.8%; Score 365; DB 16; Length 385;  
 Best Local Similarity 38.1%; Pred. No. 4.05e-31;  
 Matches 67; Conservative 54; Mismatches 49; Indels 6; Gaps 5;  
 Db 199 dfgsidiivlsia--ngpevskpplletsrkgylaaisassysfsvllsh-flpimpvgga 255  
 QY 83 DVGNIDGVYHSTAFFANMEDLRGRFSET-SREGFLAQDISYS-LIVVAHEKKLMPEGGS 141  
 Db 256 sislytiaserripiqggmksaalesdrvleagtkqniirvtisggplgraak 315  
 QY 142 IVATTIQLGGEFAVNQNNV-MGVAKASLEANVKYLAQDGP-DNIRVNAISAGPIRLSAK 199



CLASSIFICATION #superfamily ribitol dehydrogenase; short-chain alcohol dehydrogenase homology

FEATURE

7-188 #domain short-chain alcohol dehydrogenase homology

SUMMARY #length 249 #label SADH

Query Match Best Local Similarity 23.7%; Score 225; DB 5; Length 249; Pred. No. 4.0e-13; Mismatches 81; Indels 12; Gaps 12;

Matches 61; Conservative

Db 1 mnlvqdkvtitg-gt-rgigfaakifidngakvsifgetgevdtaql-kelypee 57

Qy 2 LNL-ENKYVIMTANKSIAFSAVAKVLQDGAKL-VTYKERSRKELEKILLEQINQPE 59

Db 58 evlgafapdtsrdavmaavgvqaqqkygrldmnnnacits-nnfvfrseefk-himdi 115

Qy 60 AHL-YQDYSQDEVINGFEQIGKDVGNDGTHSIAPANMDDLRGRFSETREGFLIAQ 118

D 116 nvtgvfngawcaycqcmk-dakgvintasrtvgifgs1sgvypaskasvglthlgre 174

119 DSS-XSUTIVHEAKKLMPREGGSITIVYLGEEFA/QNYNMVGAKASLEANVKYLAD 177

Db 175 lirkknirvvgapgvvntdmtgnpp-e-imegylkalpmkrmlpeefianyflfasl 232

Qy 178 LDPDNIRVNAISAGPRTLTSARGVGGNTLKEIEERAPLKRNVQDEVGRTAAYLSDL 237

Db 233 asgitatvsvdgayip 249

Qy 238 SSGVTGENIHVDGFHA 254

RESULT 8 A28053 #type complete

ENTRY

TITLE adipocyte p27 protein - mouse

ORGANISM #formal\_name Mus musculus #common\_name house mouse

DATE 19-Nov-1988 #sequence\_revision 19-Nov-1988 #text\_change 23-Mar-1995

ACCESSIONS S03382; A28053

REFERENCE

#authors Navre, M.; Ringold, G.M.

J. Cell. Biol. (1988) 107:279-286

#title A growth factor-repressible gene associated with protein kinase C-mediated inhibition of adipocyte differentiation.

#cross-references MUID:88273310

#molecule-type mRNA

#residues 1-744 #label NAV

#cross-references EMBL:X07111

CLASSIFICATION #superfamily ribitol dehydrogenase; short-chain alcohol dehydrogenase homology

FEATURE

8-180 #domain short-chain alcohol dehydrogenase homology

SUMMARY #length 244 #molecular-weight 25958 #checksum 5185

Query Match Best Local Similarity 25.3%; Score 219; DB 5; Length 244; Pred. No. 2.15e-12; Mismatches 55; Indels 4; Gaps 3;

Matches 46; Conservative

Db 61 lgwdtakalgfifgpdlvnnaalvinq-pflevtkafdrfsynrlrsfqysqmvva 119

Qy 74 INGFEQICKDVGNDGTHSIAPANMDDLRGRFSETREGFLIAQDSSSYSTIVAH-EA 132

Db 120 rdinrnrvgpsivnvvsmahnvtfnplitystsckamtlnkramamelgphkrvnsvpv 179

Qy 133 KLMPEG-GSIVATTILGGEPAVQNNVMVGAKASLEANVKYLADLGPNTRNAISA 190

Db 180 tvvldmgkvsadpefarkkerhpirkfaevdvnnsifllsrsastsgggflvda 239

Qy 191 GPIRTLSAKVGFFNTLKEIEERAPLKRNVQDEVGRTAAYLSDLSSGVTGENIHVD 250

RESULT 9 B48674 #type complete

ENTRY tropinone reductase (EC 1.1.1.236) II - jimsonweed

TITLE #formal\_name Datura stramonium #common\_name Jimsonweed, common thornapple

ORGANISM

DATE 03-May-1994 #sequence\_revision 03-May-1994 #text\_change 20-Mar-1996

ACCESSIONS B48674

REFERENCE

#authors Nakajima, K.; Hashimoto, T.; Yamada, Y.

Proc. Natl. Acad. Sci. U.S.A. (1993) 90:9591-9595

#journal #title Two tropinone reductases with different stereospecificities are short-chain dehydrogenases evolved from a common ancestor.

#accession B48674

#status preliminary

#molecule-type mRNA

#residues 1-260 #label NAK

#cross-references GB:L20474

KEYWORDS Oxidoreductase

SUMMARY #length 260 #molecular-weight 28311 #checksum 3973

Query Match Best Local Similarity 24.5%; Score 214; DB 111; Length 260; Matches 63; Conservative 82; Mismatches 97; Indels 15; Gaps 15;

Db 6 nleqtaqtg-gs-rgiqygiveelasigas-vytccsrrnqkeindctqwsrkfkgkvaa 62

Qy 3 NLENKTYYVINGIANKRSIAGVAKVLDQGAKLYT-ERERSKE-LSEKLLQEQLNQPEA 60

Db 63 svcdlssrsrgelmnvtanh-h-gklnlvnnagiykeakd-ytved-yslimsin 118

Qy 61 HLYQIDVQSD-EEVINGFPOIQGDVGNITGVYHSIAFAMMEDLRGRFSETSREGFLIAQD 119

Db 119 feasyhlsvlahpflik-asergnvvfissvsgalavpyeavygatkqandqtlrcrafew 177

Qy 120 I-SVSILTVAAHEAKKLMPEGGSITAVITYLGGEFAVQNYNMGVAKASLEANVKYLADL 178

Db 178 akdnirvngvpgpiratslvemtqdppeckenlkidrarrmgpekelamvafif 237

Qy 179 GDPNIRVNAISAGPRTL-SAKVGFFNTI-LIKEIEERAPLKRNVQDEVGRTAAYL 235

Db 236 paasyvtcgiiydggi 254

Qy 236 DLSSGVIGENIHVDGF 252

RESULT 10 A28212 #type complete

ENTRY 27-K bile acid 7-dehydroxylation protein - Eubacterium sp.

TITLE #formal\_name Eubacterium sp.

ORGANISM

DATE 30-Sep-1989 #sequence\_revision 30-Sep-1989 #text\_change 12-Apr-1995

ACCESSIONS A28212

REFERENCE

#authors Coleman, J.P.; White, W.B.; Lijewski, M.; Hylemon, P.B.

#journal J. Bacteriol. (1988) 170:2070-2077

#title Nucleotide sequence and regulation of a gene involved in bile acid 7-dehydroxylation by Eubacterium sp. strain VPI 12708.

#cross-references MUID:88137993

#molecule-type DNA

#residues 1-249 #label COL

#experimental\_source strain VPI 12708

CLASSIFICATION #superfamily ribitol dehydrogenase; short-chain alcohol dehydrogenase homology

FEATURE #domain short-chain alcohol dehydrogenase homology

**SUMMARY** #length 249 #molecular-weight 26747 #checksum 341  
**Query Match** 12.1%; Score 212; DB 5; Length 249;  
**Best Local Similarity** 23.0%; Pred. No. 1.51e-11;  
**Matches** 59; Conservative 80; Mismatches 106; Indels 12; Gaps 12;

**Db** 1 mklvqdkititg-gt-rqigfaaakflfiengakvsifgetqeetdalaql-kelype 57  
**Qy** 2 LNL-ENKTVIIMGANKRSSTAFGAKVLDLQLGAKL-VTYRKERSRKELEKUQLELNQPE 59  
**Db** 58 evlgafpditsrdayaargtvaaygrkldvminnagit-mnsvfsvseedfkn-imdi 115  
**Qy** 60 AHL-YQIDYQSDEEVINGPEQIGRDGVNIDGVYHSIAFANMELGRGESETSRGFLIAQ 118  
**Db** 116 nvngfrngawsyrcmk-dakggvintatsvggqsgqpyqggvvitlhgigre 174  
**Qy** 119 DISS-YSLTIVAHAAKKLMPGGSVIATVYLGEFAVONYNNNGVAKASLEAVKTLADL 177  
**Db** 175 iirknirvvgvapgvvdtdmtkgipp-e-iledyylk1pmkrmlkpseianvylflasd1 232  
**Qy** 178 LGPNIRVNAISAGPIRTLSAKVGFFNTILKIEERAPLKRNVDQVEVGKTAAYLLSDL 237  
**Db** 233 asgitattisvdgbyrp 249  
**Qy** 238 SSGVGENITHVDSGFHA 254

**RESULT** 11  
**ENTRY** C24706 #type complete  
**TITLE** nodG - Rhizobium meliloti plasmid  
**ALTERNATE NAMES**  
**ORGANISM** hsnC protein  
**DATE** 30-Jun-1988 #sequence\_revision 13-Jan-1995 #text\_change  
**ACCESSION** C24706; S07675; C24193  
**REFERENCE** A93638  
**authors** Debelie, F.; Sharma, S.B.  
**journal** Nucleic Acids Res. (1986) 14:7453-7472  
**title** Nucleotide sequence of Rhizobium meliloti RCR2011 genes involved in host specificity of nodulation.  
**#cross-references** PMID:87016382  
**#accession** C24706  
**#molecule-type** DNA  
**#residues** 1-245 #label DEB  
**#experimental-source** strain RCR2011 symbiotic plasmid  
**REFERENCE** S06395  
**authors** Fisher, R.F.; Swanson, J.A.; Mulligan, J.T.; Long, S.R.  
**journal** Genetics (1987) 117:191-201  
**title** Extended region of nodulation genes in Rhizobium meliloti 1021. II. Nucleotide sequence, transcription start sites and protein products.  
**accession** S07675  
**#molecule-type** DNA  
**#residues** 1-245 #label FIS  
**#experimental-source** strain 1021 symbiotic plasmid  
**REFERENCE** A94655  
**authors** Horvath, B.; Kondorosi, E.; John, M.; Schmidt, J.; Toeroek, I.; Gyoergypal, Z.; Barabas, I.; Wieneke, U.; Scheil, J.; Kondorosi, A.  
**journal** Cell (1986) 46:335-343  
**title** Organization, structure and symbiotic function of Rhizobium meliloti nodulation genes determining host specificity for alfalfa.  
**#cross-references** MVID:66272281  
**#status** C24193  
**#molecule-type** DNA  
**#cross-references** GB:M1452  
**#experimental-source** strain AK631 (a variant of strain 41)  
**#note** the sequence reported in this reference is incorrect due to multiple frameshift errors  
**COMMENT** This is one of several proteins that control host specificity of

**GENETICS** root hair infection and nodulation.  
**#gene** nodG; hsnC  
**CLASSIFICATION** #superfamily ribitol dehydrogenase; short-chain alcohol dehydrogenase homology  
**KEYWORDS** host range; nodulation  
**FEATURE** 7-183  
**#domain** short-chain alcohol dehydrogenase homology  
**label** SADH  
**length** 245 #molecular-weight 26088 #checksum 6343  
**SUMMARY** #length 245 #molecular-weight 26088 #checksum 6343  
**Query** Match 11.2%; Score 197; DB 1; Length 245;  
**Best Local Similarity** 24.1%; Pred. No. 9.25e-10;  
**Matches** 62; Conservative 73; Mismatches 108; Indels 14; Gaps 13;  
**Db** 1 mfeftrgrkalvtqasg--aiggaiararvhqqa-iv-gihgtqiek-lettlatelgdr-v 54  
**Qy** 1 MINLENKTYITMGIAINKRSAFGYAKVLDLQGAKLUVTRKERSKELKLQLQNQPPA 60  
**Db** 5 kfpanlanrdevkalgraaedlegdilvnnaqitk-dg1f1hmadpdwd ivlern1 112  
**Qy** 61 HLYQIDVQSDSEVINGFEQIGKDGNIDSYVHSIAFANMEDLRGRFSETSREGFLAQDI 120  
**Db** 113 tamfrtretq-qmrrngrinrvtsragaignpgqtncyskagningfskslageia 171  
**Qy** 121 SS-YSTIVTAHEAKKLMPGGSTVATTYVIGEPAVQNYVMYAKASLFLANVYKLADIG 179  
**Db** 172 trnivnvcapgf-e-sam-tdklnhkkekimvaiphrmgtgevavaylasdh 228  
**Qy** 180 PDNRIVNAISAGPIRTLSAKVGFFNTILKE-TEERAPLIKRNVDQVEVGKTAAYLLSDL 238  
**Db** 229 ayvtgtihvngmami 245  
**Qy** 239 SGVGENIIHVDSGFHAI 255

**RESULT** 12  
**ENTRY** A42468 #type complete  
**TITLE** 7alpha-hydroxysteroid dehydrogenase (EC 1.1.1.159) -  
**ALTERNATE NAMES** #beta-hydroxysteroid dehydrogenase (strain VPI 12708)  
**ORGANISM** #formal\_name Eubacterium sp.  
**DATE** 24-Jul-1992 #sequence\_revision 18-Sep-1992 #text\_change  
**ACCESSIONS** A42468; A36439  
**REFERENCE** A42468  
**authors** Baron, S.F.; Franklund, C.V.; Hylemon, P.B.  
**journal** J. Bacteriol. (1991) 173:4558-4569  
**title** Cloning, sequencing, and expression of the gene coding for bile acid 7alpha-hydroxysteroid dehydrogenase from Eubacterium sp. strain VPI 12708.  
**#cross-references** MVID:91310560  
**#accession** A42468  
**#status** preliminary  
**#molecule-type** DNA  
**#residues** 1-266 #label BAR  
**#cross-references** GB:M58743  
**#note** the authors translated the codon CCC for residue 95 as Phe, and CCG for residue 106 as Phe  
**REFERENCE** A36439  
**authors** Franklund, C.V.; de Prada, P.; Hylemon, P.B.  
**journal** J. Biol. Chem. (1990) 265:9842-9849  
**title** Purification and characterization of a microbial NADP-dependent bile acid 7alpha-hydroxysteroid dehydrogenase.  
**#cross-references** MVID:90277676  
**#accession** A36439  
**#status** preliminary  
**#molecule-type** protein  
**#residues** 1-22 #label FRA  
**CLASSIFICATION** #superfamily ribitol dehydrogenase; short-chain alcohol dehydrogenase homology  
**KEYWORDS** oxidoreductase



Best Local Similarity 26.9%; Pred. No. 6.78e-08;  
 Matches 71; Conservative 74; Mismatches 97; Indels 22; Gaps 20;

---

Db 7 1dgkvalrtg-ag rgigaiaavalergakvvnyahorrprkwlrs-xpmvpmlsq 63  
 Qy 4 LENRTYVIMGIANKRSIAFGVAKVLQDGAKLVETYRK-ERSRKE-LEKLLIEOLNQPEAH 61

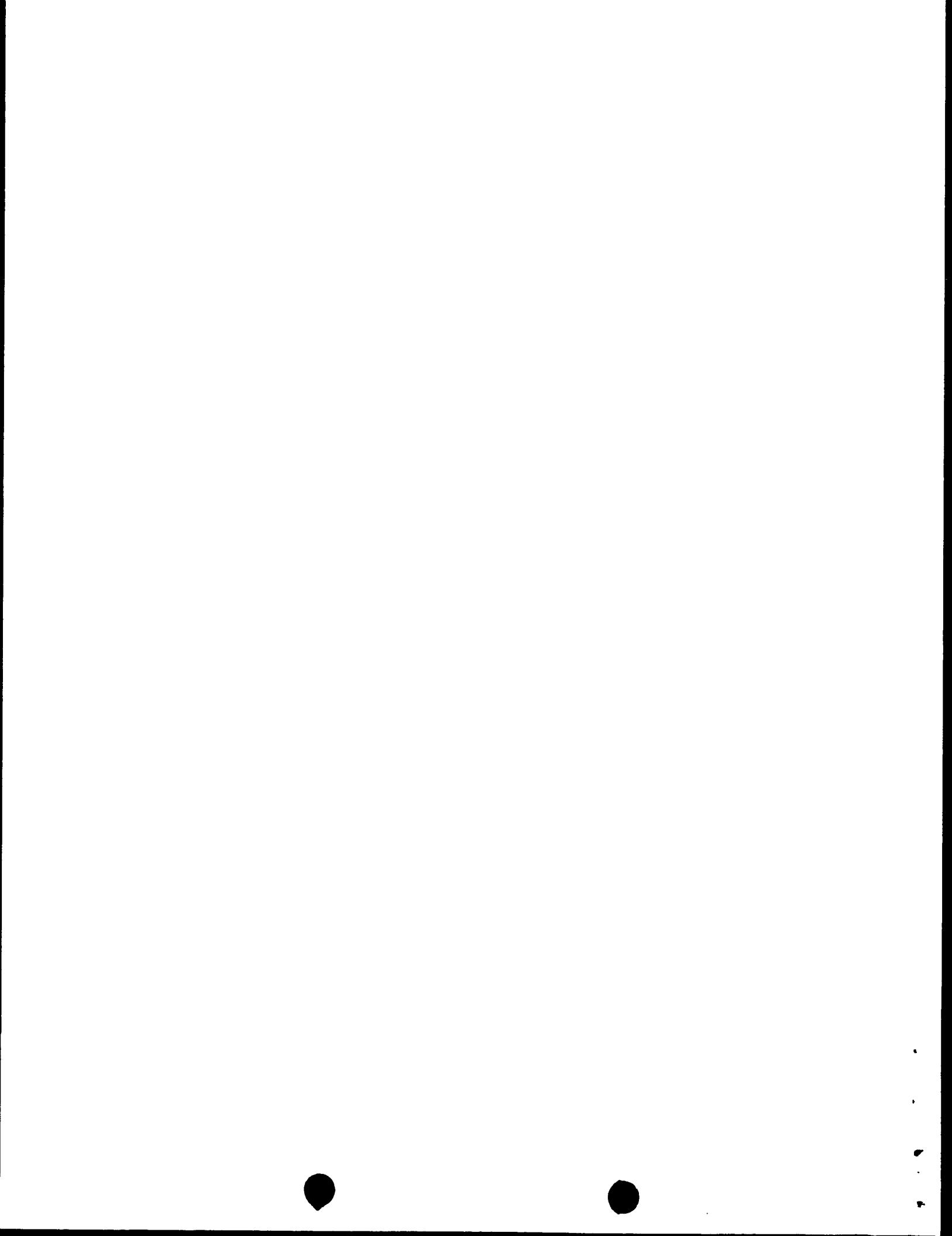
Db 64 -sgadvgdpeataklmaetvrls-s-d-agivsf-shlkdtpteeedrvfrvn 118  
 Qy 62 LYQIDVQSDBEYINGFEQIGKDGNIDGYHHSIAFANMEDLGRFSEISREGELLAQDIS 121

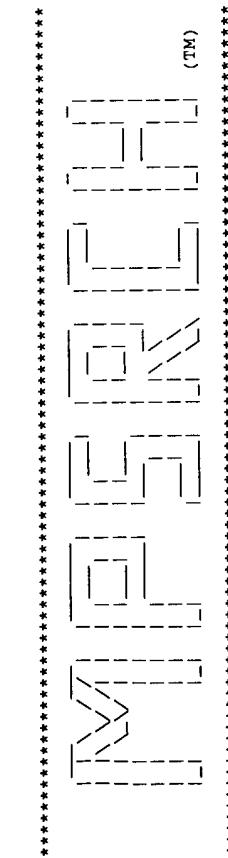
Db 119 trqqffvareayrhmreggrilitsntacvkqvpkhavysskgaidtfrxmaidcqd 178  
 Qy 122 SYSLTIVAHHEAKKLMPPEGGSIV-ATYLGEFAVQNYNNMGAKASLEANVKTLADLGP 180

Db 179 kkitvnavaapgaiktdmfлавsreyipgetftdeqydecaawisplnrvglpydvarv 238  
 Qy 181 DNIIRVNAISAGPINT---LS-AKG-VGGFNTILKE-IEBRAP-LKR-N-VD-QVEVGKTA 230

239 sfilsdtaewsgkiigvoggaf 262  
 : : | | : | : | ||: |:  
 231 AYLLSDLSSGTGENIHVDSG-FH 253

Search completed: Thu Dec 4 16:03:14 1997  
 Job time : 96 secs.





Release 2.1D John F. Collins, Biocomputing Research Unit.  
Copyright (C) 1993, 1995 University of Edinburgh, U.K.  
Distribution rights by IntelligentGenetics, Inc.

Mpsrch\_nn n.a. - n.a. database search, using Smith-Waterman algorithm

Date: Fri Dec 5 08:13:12 1997: MasPar time 4.49 Seconds  
759,327 Million cell updates/sec  
Regular output not generated.

Title: >US-08-790-043A-2  
Description: (1-771) from US58790043A.seq  
Perfect Score: 771

N.A. Sequence: 1 ATGTTAAATCTTGGAAACAA.....GATTCAGGCAATTAAATAA 771  
Comp: TACATTAGAACCTTGT.....CTAAGGTCGTTAATTATT

Scoring table:  
TABLE default  
Gap 6

Match STD : Dbase 0; Query 0  
Searched: 8249 seqs, 21907491 bases x 2  
Post-processing: Minimum Match 0<sup>a</sup>  
Listing first 45 summaries

Database: n-issued  
1:back1 2:51 3:52 4:53 5:54 6:55 7:56 8:PCT90 9:PCT91  
10:PCT92 11:PCT93 12:PCT94 13:PCT95 14:PCT96

Statistics: Mean 8.114; Variance 4.451; scale 1.823

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

### SUMMARIES

Result No.	Score	Query	Match	Length	DB ID	Description	Pred. No.
c 1	55	7..1	7218	7	US-08-232-	Sequence 14, Applicati	4..3e-22
c 2	27	3..5	215	6	US-08-238-	Sequence 5, Applicatio	4.14e-04
c 3	27	3..5	242	7	US-08-273-	Sequence 1, Applicatio	4.14e-04
c 4	24	3..1	215	6	US-08-238-	Sequence 5, Applicati	2.03e-02
c 5	22	2..9	2894	7	US-08-483-	Sequence 1, Applicatio	2.43e-01
c 6	22	2..9	2894	7	US-08-472-	Sequence 1, Applicatio	2.43e-01
c 7	22	2..9	2894	6	US-08-278-	Sequence 1, Applicatio	2.43e-01
c 8	21	2..7	74	13	PCT-US95..1	Sequence 100, Applicat	8.12e-01
c 9	21	2..7	74	13	PCT-US95..1	Sequence 94, Applicati	8.12e-01
c 10	21	2..7	81	13	PCT-US95..1	Sequence 98, Applicati	8.12e-01
c 11	21	2..7	81	13	PCT-US95..1	Sequence 92, Applicati	8.12e-01
c 12	21	2..7	242	7	US-08-273-	Sequence 1, Applicatio	8.12e-01
c 13	21	2..7	774	9	PCT-US91..0	Sequence 1, Applicatio	8.12e-01
c 14	21	2..7	824	7	US-08-158-	Sequence 1, Applicatio	8.12e-01
c 15	20	2..6	75	13	PCT-US95..1	Sequence 99, Applicati	2.63e+00
c 16	20	2..6	81	13	PCT-US95..1	Sequence 98, Applicati	2.63e+00
c 17	20	2..6	82	13	PCT-US95..1	Sequence 97, Applicati	2.63e+00
c 18	20	2..6	579	14	PCT-US96..0	Sequence 1654, Applica	2.63e+00
c 19	20	2..6	1265	7	US-08-182-	Sequence 5, Applicatio	2.63e+00

### ALIGNMENTS

RESULT 1  
ID US-08-232-463-14 STANDARD; DNA; UNC; 7218 EP.

AC xxxxx

DT 01-JAN-1900  
DE Sequence 14, Application US/08232463.  
CC Sequence 14, Application US/08232463  
Patent No. 5670367  
GENERAL INFORMATION:  
CC APPLICANT: DORNER, F.  
CC APPLICANT: SCHEIFFLINGER, F.  
CC APPLICANT: FALKNER, F. G.  
TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS  
NUMBER OF SEQUENCES: 52  
CORRESPONDENCE ADDRESS:  
CC ADDRESSEE: Foley & Lardner  
STREET: 1800 Diagonal Road, Suite 500  
CITY: Alexandria  
STATE: VA  
COUNTRY: USA  
ZIP: 22313-0299

COMPUTER READABLE FORM:  
MEDIUM TYPE: FLOPPY disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/232,463  
FILING DATE:  
CLASSIFICATION: 435  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: US/07/935,313  
FILING DATE:  
APPLICATION NUMBER: EP 91 114 300 .6  
FILING DATE: 26-AUG-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: BENT, Stephen A.  
REGISTRATION NUMBER: 29,768  
REFERENCE/DOCKET NUMBER: 30472/114 IMMU  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703)836-9300  
TELEFAX: (703)83-4109  
TELEX: 899149  
INFORMATION FOR SEQ ID NO: 14:





CP 450 GCCACCTAAATATGTTGGCAACAAATGCTTACCACTTCCTGGCATTAAT 402

**RESULT** 5 ID US-08-483-859-1 STANDARD; DNA; UNC; 2894 BP.  
 AC xxxx DT 01-JAN-1900 DE Sequence 1, Application US/08483859.  
 CC Patent No. 5656436  
 CC GENERAL INFORMATION:  
 CC APPLICANT: LOOSMORE, Sheena M.  
 CC APPLICANT: YANG, Yan-Ping  
 CC APPLICANT: CHONG, Pele  
 CC APPLICANT: OOMEN, Raymond P.  
 CC APPLICANT: KLEIN, Michel H.  
 CC TITLE OF INVENTION: Analog of Haemophilus Hin47 Protein with  
 CC TITLE OF INVENTION: Reduced Protease Activity  
 CC NUMBER OF SEQUENCES: 23  
 CC CORRESPONDENCE ADDRESS:  
 CC ADDRESSEE: Sim & McBurney  
 CC STREET: Suite 701, 330 University Avenue  
 CC CITY: Toronto  
 CC STATE: Ontario  
 CC COUNTRY: Canada  
 CC ZIP: M5G 1R7  
 CC COMPUTER READABLE FORM:  
 CC MEDIUM TYPE: FLOPPY disk  
 CC COMPUTER: IBM PC compatible  
 CC OPERATING SYSTEM: PC-DOS/MS-DOS  
 CC SOFTWARE: PatentIn Release #1.0, version #1.25  
 CC CURRENT APPLICATION DATA:  
 CC APPLICATION NUMBER: US 08/296,149  
 CC FILING DATE: 07-JUN-1995  
 CC CLASSIFICATION: 435  
 CC PRIOR APPLICATION DATA:  
 CC APPLICATION NUMBER: US 08/296,149  
 CC FILING DATE: 26-AUG-1994  
 CC PRIORITY APPLICATION DATA:  
 CC APPLICATION NUMBER: US 08/278,091  
 CC FILING DATE: 07-JUN-1994  
 CC CLASSIFICATION: 435  
 CC ATTORNEY/AGENT INFORMATION:  
 CC NAME: Stewart, Michael I.  
 CC REGISTRATION NUMBER: 24,973  
 CC REFERENCE/DOCKET NUMBER: 1038-493 MIS:vg  
 CC TELECOMMUNICATION INFORMATION:  
 CC TELEPHONE: (416) 595-1155  
 CC TELEFAX: (416) 595-1163  
 CC INFORMATION FOR SEQ ID NO: 1:  
 CC SEQUENCE CHARACTERISTICS:  
 CC LENGTH: 2894 base pairs  
 CC TYPE: nucleic acid  
 CC STRANDEDNESS: single  
 CC TOPOLOGY: linear  
 SQ Sequence 2894 BP; 897 A; 570 C; 515 G; 912 T; 0 other;  
 SQ Query Match 7 ID US-08-278-091-1 STANDARD; DNA; UNC; 2894 BP.  
 AC xxxx AC 2.9%; Score 22; DB 7; Length 2894;  
 CC Best Local Similarity 75.0%; Pred. No. 2.43e-01;  
 CC Matches 33; Conservative 0; Mismatches 11; Indels 0; Gaps 0;  
 CC DE Sequence 1, Application US/08278091.  
 CC Sequence 1, Application US/08278091.  
 CC Patent No. 5566139  
 CC GENERAL INFORMATION:  
 CC APPLICANT: LOOSMORE, Sheena M.  
 CC APPLICANT: YANG, Yan-Ping  
 CC APPLICANT: CHONG, Pele  
 CC APPLICANT: OOMEN, Raymond P.  
 CC APPLICANT: KLEIN, Michel H.  
 CC TITLE OF INVENTION: Analog of Haemophilus Hin47 Protein with  
 CC TITLE OF INVENTION: Reduced Protease Activity  
 CC NUMBER OF SEQUENCES: 23  
 CC CORRESPONDENCE ADDRESS:  
 CC ADDRESSEE: Sim & McBurney  
 CC STREET: Suite 701, 330 University Avenue  
 CC

CITY: Toronto  
 STATE: Ontario  
 COUNTRY: Canada  
 ZIP: M5G 1R7

COMPUTER READABLE FORM:  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patentin Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/278,091  
 FILING DATE: 21-JUL-1994  
 CLASSIFICATION: 435  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Stewart, Michael T  
 REGISTRATION NUMBER: 24,973  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (416) 595-1155  
 TELEFAX: (416) 595-1153  
 LENGTH: 2894 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear

SEQUENCE CHARACTERISTICS:  
 Sequence 2894 BP; 897 A; 570 C; 515 G; 912 T; 0 other;  
 Query Match Score 22; DB 6; Length 2894;  
 Best Local Similarity 75.0%; Pred. No. 2.43e-01;  
 Matches 33; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

RESULT 8 PCT-US95-11934-100 STANDARD: DNA; UNC: 74 BP.  
 ID PCT-US95-11934-100  
 DT 01-JAN-1990  
 DE Sequence 100, Application PC/TUS9511934.  
 SEQUENCE 100, Application PC/TUS9511934  
 GENERAL INFORMATION:  
 CC  
 APPLICANT: Cyrogen Corporation  
 TITLE OF INVENTION: Antigen Binding Peptides (Abtides) From  
 Peptide Libraries  
 NUMBER OF SEQUENCES: 103  
 CORRESPONDENCE ADDRESS:  
 STREET: 1155 Avenue of the Americas  
 CITY: New York  
 STATE: New York  
 COUNTRY: USA  
 CC  
 ZIP: 10036  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: FLOPPY disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patentin Release #1.0, Version #1.30  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: PCT-US95/11934  
 FILING DATE: 20-SEP-1995  
 CLASSIFICATION:  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Misrock, S. Leslie  
 REGISTRATION NUMBER: 18,872  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (212) 750-9090  
 TELEFAX: (212) 869-9741/8864  
 TELEX: 66141 PENNIE  
 INFORMATION FOR SEQ ID NO: 94:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 74 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: DNA (genomic)  
 Sequence 74 BP; 6 A; 6 C; 1 G; 1 T; 60 other;  
 Query Match Score 21; DB 13; Length 74;  
 Best Local Similarity 4.8%; Pred. No. 8.12e-01;  
 Matches 3; Conservative 19; Mismatches 41; Indels 0; Gaps 0;

RESULT 9 PCT-US95-11934-94 STANDARD: DNA; UNC: 74 BP.  
 ID PCT-US95-11934-94  
 AC XXXXXX  
 DT 01-JAN-1990  
 DE Sequence 94, Application PC/TUS9511934.  
 SEQUENCE 94, Application PC/TUS9511934  
 GENERAL INFORMATION:  
 CC  
 APPLICANT: Cyrogen Corporation  
 TITLE OF INVENTION: Antigen Binding Peptides (Abtides) From  
 Peptide Libraries  
 NUMBER OF SEQUENCES: 103  
 CORRESPONDENCE ADDRESS:  
 CC  
 STREET: 1155 Avenue of the Americas  
 CITY: New York  
 STATE: New York  
 COUNTRY: USA  
 CC  
 ZIP: 10036  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: FLOPPY disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patentin Release #1.0, Version #1.30  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: PCT-US95/11934  
 FILING DATE: 20-SEP-1995  
 CLASSIFICATION:  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Misrock, S. Leslie  
 REGISTRATION NUMBER: 18,872  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (212) 750-9090  
 TELEFAX: (212) 869-9741/8864  
 TELEX: 66141 PENNIE  
 INFORMATION FOR SEQ ID NO: 95:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 74 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: DNA (genomic)  
 Sequence 74 BP; 3 A; 4 C; 3 G; 1 T; 63 other;  
 Query Match Score 21; DB 13; Length 74;  
 Best Local Similarity 7.2%; Pred. No. 8.12e-01;  
 Matches 5; Conservative 19; Mismatches 45; Indels 0; Gaps 0;

RESULT 10 PCT-US95-11934-101 STANDARD: DNA; UNC: 74 BP.  
 ID PCT-US95-11934-101  
 DT 01-JAN-1990  
 DE Sequence 101, Application PC/TUS9511934.  
 SEQUENCE 101, Application PC/TUS9511934  
 GENERAL INFORMATION:  
 CC  
 APPLICANT: Cyrogen Corporation  
 TITLE OF INVENTION: Antigen Binding Peptides (Abtides) From  
 Peptide Libraries  
 NUMBER OF SEQUENCES: 103  
 CORRESPONDENCE ADDRESS:  
 ADDRESSSEE: Pennie & Edmonds  
 STREET: 1155 Avenue of the Americas  
 CITY: New York  
 STATE: New York  
 COUNTRY: USA  
 ZIP: 10036  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: FLOPPY disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patentin Release #1.0, Version #1.30  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: PCT-US95/11934  
 FILING DATE: 20-SEP-1995  
 CLASSIFICATION:  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Misrock, S. Leslie  
 REGISTRATION NUMBER: 18,872  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (212) 750-9090  
 TELEFAX: (212) 869-9741/8864  
 TELEX: 66141 PENNIE  
 INFORMATION FOR SEQ ID NO: 100:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 74 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: DNA (genomic)  
 Sequence 74 BP; 3 A; 4 C; 3 G; 1 T; 63 other;  
 Query Match Score 21; DB 13; Length 74;  
 Best Local Similarity 7.2%; Pred. No. 8.12e-01;  
 Matches 5; Conservative 19; Mismatches 45; Indels 0; Gaps 0;

RESULT 11 PCT-US95-11934-102 STANDARD: DNA; UNC: 74 BP.  
 ID PCT-US95-11934-102  
 DT 01-JAN-1990  
 DE Sequence 102, Application PC/TUS9511934.  
 SEQUENCE 102, Application PC/TUS9511934  
 GENERAL INFORMATION:  
 CC  
 APPLICANT: Cyrogen Corporation  
 TITLE OF INVENTION: Antigen Binding Peptides (Abtides) From  
 Peptide Libraries  
 NUMBER OF SEQUENCES: 103  
 CORRESPONDENCE ADDRESS:  
 ADDRESSSEE: Pennie & Edmonds  
 STREET: 1155 Avenue of the Americas  
 CITY: New York  
 STATE: New York  
 COUNTRY: USA  
 ZIP: 10036  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: FLOPPY disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patentin Release #1.0, Version #1.30  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: PCT-US95/11934  
 FILING DATE: 20-SEP-1995  
 CLASSIFICATION:  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Misrock, S. Leslie  
 REGISTRATION NUMBER: 18,872  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (212) 750-9090  
 TELEFAX: (212) 869-9741/8864  
 TELEX: 66141 PENNIE  
 INFORMATION FOR SEQ ID NO: 100:





Fri Dec 5 12:31:36 1997

US-08-790-043A-2.rni

Page 8

DE Sequence 1, Application US/08158353.  
CC Sequence 1, Application US/08158353.  
CC Patent No. 5620862  
GENERAL INFORMATION:  
CC APPLICANT: Padula, Steven J.  
CC TITLE OF INVENTION: Methods for Diagnosing Early Lyme  
CC TITLE OF INVENTION: Disease  
CC NUMBER OF SEQUENCES: 7  
CC CORRESPONDENCE ADDRESS:  
CC ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.  
CC STREET: Two Militia Drive  
CC CITY: Lexington  
CC STATE: MA  
CC COUNTRY: USA  
CC ZIP: 02173  
CC COMPUTER READABLE FORM:  
CC MEDIUM TYPE: Floppy disk  
CC COMPUTER: IBM PC compatible  
CC OPERATING SYSTEM: PC-DOS/MS-DOS  
CC SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
CC APPLICATION NUMBER: US/08/158,353  
CC FILING DATE:  
CC CLASSIFICATION: 435  
CC ATTORNEY/AGENT INFORMATION:  
CC NAME: Carroll, Alice O.  
CC REGISTRATION NUMBER: 33,542  
CC REFERENCE/DOCKET NUMBER: UCT93-05  
TELECOMMUNICATION INFORMATION:  
CC TELEFAX: 617-861-6240  
CC INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
CC LENGTH: 824 base Pairs  
CC TYPE: nucleic acid  
CC STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
Sequence 824 BP; 339 A; 104 C; 143 G; 238 T; 0 other;  
  
Query Match 2.7%; Score 21; DB 7; Length 824;  
Best Local Similarity 83.9%; Pred. No. 8.12e-01;  
Matches 26; Conservative 0; Mismatches 5; Indels 0; Gaps 0;  
  
Db 703 AAAGCTTGAAGATTGATCAGTAA 733  
||| ||| ||| ||| ||| ||| ||| ||| |||  
QY 137 AAGAGCTTGAATAATTAGACATAAA 167  
||| ||| ||| ||| ||| ||| ||| |||  
T 15  
1B PCT-US95-11934-99 STANDARD; DNA; UNC; 75 BP.  
AC xxxxxx  
DT 01-JAN-1900  
DE Sequence 99, Application PC/TUS9511934.  
CC Sequence 99, Application PC/TUS9511934.  
GENERAL INFORMATION:  
CC APPLICANT: Cyrogen Corporation  
TITLE OF INVENTION: Antigen Binding Peptides (Abtides) From  
CC NUMBER OF SEQUENCES: 103  
CORRESPONDENCE ADDRESS:  
CC CC ADDRESSEE: Pennie & Edmonds  
STREET: 1155 Avenue of the Americas  
CC CITY: New York  
STATE: New York  
CC COUNTRY: USA  
ZIP: 10036  
CC COMPUTER READABLE FORM:  
CC MEDIUM TYPE: Floppy disk  
CC COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:

Release 2.1D John F. Collins Biocomputing Research Unit.  
Copyright (c) 1993-1994, 1995 University of Edinburgh, U.K.  
Distribution rights by IntelliGenetics, Inc.

psrch\_nn n.a. - n.a. database search, using Smith-Waterman algorithm

on:	Fri Dec 5 07:57:33 1997;	MasPar time 100.32 Seconds
		802.071 Million cell updates/sec

Lar output not generated.

Title:	>US-08-790-043A-2		
Description:	(1-771) from US08790043A.seq		
Perfect Score:	771	A.	Sequence:
Comp.:			TACATTAGAACTTGTTCATTTTCTAAGGNGCGTAAATTATT
Scoring Table:	TABLE default		

match STD : Dbase 0; Query 0

```
host-processing: Minimum Match 0%
Listing first 45 summaries

n-geoseq28
 1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7
 8:part8 9:part9 10:part10 11:part11 12:part12 13:part13
 14:part14 15:part15 16:part16 17:part17 18:part18
 19:part19 20:part20 21:part21 22:part22 23:part23
 24:part24 25:part25 26:part26 27:part27 28:part28
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

result	No.	Score	Query No.	Match	Length	DB ID	Description	Pred. No.
1	63	8.2	1047	2	Q10572	Human Natriuretic Pep	3.58E-23	
2	52	6.7	1047	2	Q10572	Human Natriuretic Pep	1.56E-16	
3	40	5.2	91	9	Q51746	Oligonucleotide probe	1.42E-09	
4	38	4.9	204	1	N81164	Base substituted E.co	1.86E-08	
5	37	4.8	91	9	N811746	Oligonucleotide probe	6.67E-08	
6	34	4.4	204	1	N81164	Base substituted E.co	9.99E-06	
7	33	4.3	114	12	Q70467	Generic DNA sequence	3.41E-05	
8	32	4.2	114	12	Q70465	Generic DNA sequence	1.15E-04	
9	31	4.0	114	12	Q70469	Generic DNA sequence	3.81E-04	
10	30	3.9	114	12	Q70468	Generic DNA sequence	3.81E-04	
11	30	3.9	114	12	Q70466	Generic DNA sequence	4.05E-03	
12	28	3.6	114	12	Q70470	Generic DNA sequence	1.25E-01	
13	25	3.2	114	12	Q70465	Generic DNA sequence	3.76E-01	
14	24	3.1	114	12	Q70470	Generic DNA sequence	3.76E-01	
15	24	3.1	114	12	Q70469	Generic DNA sequence	3.76E-01	

c	39	21	2.7	81	21	T13611
	40	21	2.7	453	8	G9418
	41	21	2.7	501	3	N50023
	42	21	2.7	534	12	Q73883
	43	21	2.7	774	3	Q22883
	44	21	2.7	2403	21	T12562
	45	21	2.7	5746	24	T09225

ALIGNMENTS

Q10572;	standard; DNA; 104 / BP.
Q10572;	standard; DNA; 104 / BP.
09 APR-1991 (first entry)	
Human Natriuretic Peptide Receptor	
NPBP; ANP; BNP; CNP; Kidney failure	
hyperaldosteronism; glaucoma; guaa-	
nos	
Homo sapiens	
Key	Location/Qualifie
PFH	1..22
label= signal sequence	
Protein	12
PFH	versus
FFT	
PFH	

Domain	23...455
/label=	extracellular domain
/note-	"binds natriuretic peptide
Domain	456...456
/label=	transmembrane domain
Domain	479...1047
/label=	cytoplasmic domain
/note-	"GC and protein kinase act
Modified -site	24...26
/label=	N-glycosyl
Modified -site	35...37
/label=	N-glycosyl
Modified -site	161...163
/label=	N-glycosyl
Modified -site	195...197
/label=	N-glycosyl
Modified -site	244...246
/label=	N-glycosyl
Modified -site	277...279
/label=	N-glycosyl
Modified -site	349...351
/label=	N-glycosyl
Modified -site	600...602
/label=	N-glycosyl
	W0100192-A.
	10-TAN-1991
FT	PD

heart failure; protein kinase C; cyclase.





PR	31-JAN-1994;	US-189331.
	(UYNC-)	UNIV NORTH CAROLINA.
ID	N81164 standard; DNA;	204 BP.
AC	N81164;	
DT	08-NOV-1990 (first entry)	
DE	Base substituted E.coli beta-galactosidase alpha-fragment; base substitutions; ss.	
KW	E.coli beta galactosidase alpha-fragment; base substitutions; ss.	
OS	Escherichia coli.	
Key	Location/Qualifiers	
misc-feature	19..69	
FT	/tag= a	
FT	/function=multiple cloning site	
FT	primer_bind 187..204	
FT	/*tag= b	
PN	EP-285123-A.	
PD	05-MAY-1988.	
PT	Introducing random point mutations into nucleic acids -	
PT	by prep of single stranded template, annealing a primer, elongation, misincorporation, completion of molecules and screening.	
PS	Disclosure: P: English.	
CC	Random point mutations were introduced into the alpha fragment of E.coli beta-galactosidase. The wild type sequence was obtained as a single stranded template and an Oligonucleotide was hybridised to it to generate a pool of DNA molecules which terminate at all possible nucleotide positions within a specified region. The variable 3' ends generated in this way are used as primers for reverse transcriptase. Nucleotides are misincorporated by the transcriptase and the molecules are completed to forms that can be amplified and then expressed in a suitable host-vector system.	
CC	The sequence covers all 176 diffet base substitutions, most of which occurred singularly in any given mutant.	
CC	See also P8075.	
Sequence	204 BP; 21 A; 47 C; 17 G; 11 T; 108 Others;	
Query Match	Score 34; DB 1; Length 204;	
Best Local Similarity	4.4%	
Matches	54; Mismatches 41; Indels 0; Gaps 0;	
Db	84 cgtmrttthhyrrmrbrnvyrdynrsdaawyccyrrsvkyyddcynochhhhyvbbv 143	
Qy	42 CCCTATAAAGCCTAGTATTGCTTTAGATTTAGATCAATTAGGTCTAA 101	
144 nvhnhnnncnnccbnvchnvhbnhrwayrhdaarrdhv 185		
Query	102 ATTAGTATTACTTACCGTAAAGAACGTAGCCGTTAAAGAGCT 143	
RESULT	7	
ID	Q70467 standard; DNA; 114 BP.	
AC	Q70467;	
DT	05-APR-1995 (first entry)	
DE	Generic DNA sequence to generate a random TSAR peptide library.	
KW	TSAR; totally synthetic affinity reagent; synthetic; binding domain; effector domain; concatenated heterofunctional protein; linker;	
KW	direct; rapid; detection; screening; treatment; generic; ss.	
Synthetic		
Key	Location/Qualifiers	
misc-feature	55..60	
FT	/*tag= a	
FT	/note= "this sequence represents 'z'; z can be a sequence of 6, 9 or 12 nucleotides (see comments)"	
FT	W09418318 A.	
FT	18-AUG-1994.	
FT	000977.	
FT	01-FEB-1994; W00977.	
FT	01-FEB-1994; US-011416.	
FT	01-FEB-1993; US-17500.	
FT	30-DEC-1993; US-17500.	
FT	31-JAN-1994; US-18331.	
PA	(UYNC-)	UNIV NORTH CAROLINA.
PD	18-AUG-1994.	
PF	01-FEB-1994; W00977.	
PR	01-FEB-1993; US-011416.	
DR	WPI; 94-27973/34.	
P-PDB	R65150 and R65151.	
PT	Identifying Proteins or Peptide(s) which bind a ligand - by screening a recombinant vector library expressing fusion proteins	
PR	31-JAN-1994; US-189331.	
AC	Q70465;	
DT	05-APR-1995 (first entry)	
DE	Generic DNA sequence to generate a random TSAR peptide library.	
KW	TSAR; totally synthetic affinity reagent; synthetic; binding domain; effector domain; concatenated heterofunctional protein; linker;	
KW	direct; rapid; detection; screening; treatment; generic; ss.	
Synthetic		
Key	Location/Qualifiers	
misc-feature	55..60	
FT	/*tag= a	
FT	/note= "this sequence represents 'z'; z can be a sequence of 6, 9 or 12 nucleotides (see comments)"	
FT	W09418318 A.	
FT	18-AUG-1994.	
FT	000977.	
FT	01-FEB-1994; W00977.	
FT	01-FEB-1993; US-011416.	
FT	01-FEB-1993; US-17500.	
FT	30-DEC-1993; US-17500.	
FT	31-JAN-1994; US-18331.	
PA	(UYNC-)	UNIV NORTH CAROLINA.
PD	18-AUG-1994.	
PF	01-FEB-1994; W00977.	
PR	01-FEB-1993; US-011416.	
DR	WPI; 94-27973/34.	
P-PDB	R65150 and R65151.	
PT	Identifying Proteins or Peptide(s) which bind a ligand - by screening a recombinant vector library expressing fusion proteins	

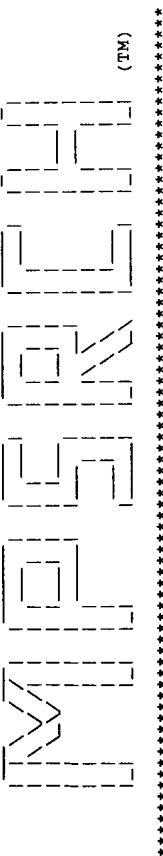
comprising a binding domain and an effector domain Disclosure; Page 35; 255pp; English.

Q70465 is a generic DNA sequence used to generate random TSAR (Totally Synthetic Affinity Reagents) peptides. This generic formula can also be represented as follows: X(NNB)6(TGC)(NNB)11Z(NNB)14(TGC)(NNB)3Y. X and Y are flanking restriction sites (X is not the same as Y) that are not specified further. Other generic sequences are shown in Q70466-68. Other specific peptides generated by these generic sequences are shown in R65151-54. TSARs are concatenated heterofunctional proteins or peptides, comprising at least two functional regions - a binding domain with affinity for a ligand and a second effector peptide portion that is chemically or biologically active. They may further comprise a linker peptide between the 2 domains. The oligonucleotides are also designed so that the expressed peptide contains 2 or 4 cysteine residues positioned in, or flanking, the unpredicted or variant residues. These residues confer some degree of conformational rigidity to the peptides. The TSARs or compsns. comprising a TSAR binding domain can be used *in vivo* to deliver a chemically or biologically active moiety, eg. metal ion, radioisotope, peptide, toxin or enzyme, to the specific target or on the cell. They can also replace the function of macromolecules, eg. monoclonal or polyclonal antibodies and therefore circumvent the need for complex methods of hybridoma formation or *in vivo* antibody production. The TSARs are easily characterised and have designed activity allowing direct and rapid detection in a screening process.







\*\*\*\*\*  
  
\*\*\*\*\*

Release 2.1D John F. Collins, Biocomputing Research Unit, University of Edinburgh, U.K.

Copyright (c) 1993, 1994, 1995 University of Edinburgh, Inc.

Distribution rights by IntelliGenetics, Inc.

n.a. - n.a. database search, using Smith-Waterman algorithm

Fri Dec 5 07:44:28 1997; MasPar time 750.09 Seconds  
1128.898 Million cell updates/sec

Similar output not generated.

Title: >US-08-790-043A-2

Description: (1-71) from US08790043A.seq

Perfect Score: N/A

Sequence: Comp: ATGTTAAATCTGTAAAACAA.....GATTCCAGCAATTAAATAAA 771  
TACAATTAGAACTTTCTT.....CTAAGGGCGTTAAATTATT

Scoring table: TABLE default  
Gap 6

Nmatch STD : Dbase 0; Query 0

Searched: 362067 seqs, 549138275 bases x 2

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: embl-new3

1:BCT 2:FUN 3:GEN1 4:GEN2 5:HTG1 6:HTG2 7:HUM 8:INV  
9:ORG 10:MAM 11:VRT 12:PLN 13:PRO 14:ROD 15:SYN 16:UNC  
17:YTR

Database: genbank99

18:BCT1 19:BCT2 20:BCT3 21:BCT4 22:BCT5 23:BCT6 24:BCT7  
25:RCT8 26:RCT9 27:BCT10 28:BCT11 29:SENI 30:GEN2

31:GEN3 32:HG1 33:HTG2 34:HTG3 35:INV1 36:INV2 37:INV3  
38:INV4 39:INV5 40:INV6 41:INV7 42:INV8 43:INV9 44:INV10  
45:INV11 46:INV12 47:MAIM2 48:MAIM3 49:VPT1 50:VRT2 51:VRT3  
52:VRT4 53:VRT1 54:VRT2 55:VRT3 56:VRT4 57:PATS 58:PIG

59:PIUN1 60:PLN1 61:PLN3 62:PLN4 63:PLN5 64:PLN6 65:PLN7  
66:PLN8 67:PLN9 68:PLN10 69:PLN11 70:PLN12 71:PRI2  
72:PRI3 73:PRI4 74:PRI5 75:PRI6 76:PRI7 77:PRI8 78:PRI9  
79:PRI10 80:PRI11 81:PRI12 82:PRI13 83:PRI14 84:PRI15  
85:ROD1 86:ROD2 87:ROD3 88:ROD4 89:ROD5 90:ROD6 91:ROD7  
92:ROD8 93:STR 94:SYN 95:UNA 96:VRL1 97:VRL2 98:VRL3  
99:VRL4 100:VRL5 101:VRL6 102:VRL7 103:VRL8 104:VRL9  
105:VRL10

genbank-new3

106:BCT 107:GEN1 108:GEN2 109:HTG1 110:HTG2 111:INV  
112:NAME 113:VRT 114:PHG 115:PIN 116:PRI1 117:PRI2  
118:ROD 119:SYN 120:UNA 121:VRL  
u-emb150\_99

Database: Statistics: Mean 10.783; Variance 5.167; scale 2.087

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

JOURNAL Source 93308081

MEDLINE /tissue lib="PCC 7120"

FEATURES /organism="Anabaena sp."

ORGANISM "Anabaena sp."

LOCATION/QUALIFIERS 1..4142

KEYWORDS DNA binding protein; bfa gene.

AUTHORS Wei,T.-F., Ramasubramanian,T.S., Pu,F., and Golden,J.W.

ORGANISM Anabaena sp. (Library: PCC 7120) DNA.

DEFINITION Eubacteria; Cyanobacteria; Nostocales; Nostocaceae; Anabaena.

ACCESSION L10036

REFERENCE 9142008

ARTICLE 1 (bases 1 to 4142)

SELECTOR Anabaena sp. strain PCC 7120 bfa gene encoding a sequence-specific selection

JOURNAL J. Bacteriol. 175, 4025-4035 (1993)

SUMMARIES

	Result No.	Score	Query Length	DB ID	Description	Pred. No.
	1	110	14.3	4142 18	Anabaena sp. sequence	3.86e-64
	2	89	11.5	122056 20	Synechocystis sp. PCC 4.50e-47	
	3	66	8.6	2123 25	Proteus mirabilis glu 5.34e-29	
	C	4	56	7.3	Hil32846 1.87e-21	
	C	5	56	7.3	Haemophilus influenza 1.87e-21	
	C	6	43	5.6	Hil3246 4.14e-12	
	(TM)				Escherichia coli shor	4.14e-12
					E. coli envM gene.	4.14e-12
					Escherichia coli from	4.14e-12
					E. coli genomic DNA, K	4.14e-12
					E. coli genomic DNA, K	4.14e-12
					E. coli genomic DNA, K	4.14e-12
					E. coli genomic DNA, K	4.14e-12
					S. typhimurium envM pr	4.66e-10
					Oxytricha fallax 57KD 1.26e-03	
					Oxytricha fallax 57KD 1.26e-03	
					Mycoplasma bovine gro 1.26e-03	
					Mycoplasma mycooides g 1.9e-02	
					Mycoplasma mycooides g 1.9e-02	
					Mycoplasma mycooides g 1.9e-02	
					Sequence 5 from paten	7.2le-02
					Acinetobacter calcoac 7.2le-02	
					A. calcoaceticus quino 7.21e-02	
					Helicoverpa armigera 2.64e-02	
					D. discoideum mRNA fr 9.38e-01	
					Wh1l-switch-regulated 9.38e-01	
					Slime mold (D. burgundy 9.38e-01	
					S. cerevisiae chromoso 9.38e-01	
					S. cerevisiae SPX19 an 9.38e-01	
					B. burgdorferi 9.38e-01	
					H. sapiens LAV3/BCL16 9.38e-01	
					S. dysgalactiae fnbB g 9.38e-01	
					S. cerevisiae DNA of c 9.38e-01	
					Sequence 5 from paten 3.2e-00	
					Human 33-kDa phototra 3.23e+00	
					Mycoplasma capricolum 3.23e+00	
					Caenorhabditis elegan 3.23e+00	
					Toxoplasma gondii chl 1.07e+01	
					Synchocystis sp. PCC 1.07e+01	
					Clostridium acetobuty 1.07e+01	
					S. equisimilis (Se165) 1.07e+01	
					Schistosoma mansoni m 1.07e+01	
					Caenorhabditis elegan 1.07e+01	
					Toxoplasma gondii chl 1.07e+01	
					Synchocystis sp. PCC 1.07e+01	
					Clostridium acetobuty 3.44e+01	
					B. aphidicola rpII, rp 3.44e+01	
					Caenorhabditis elegan 3.44e+01	
					ALIGNMENTS	
					RESULT 1	
					LOCUS ANAB1FA 4142 bp	
					DEFINITION DNA sequence-specific DNA binding protein (bfa) gene,	
					ORGANISM Anabaena sp. (Library: PCC 7120) DNA.	
					ACCESSION L10036	
					NID 9142008	
					KEYWORDS DNA binding protein; bfa gene.	
					AUTHORS Wei,T.-F., Ramasubramanian,T.S., Pu,F., and Golden,J.W.	
					ORGANISM Anabaena sp. strain PCC 7120 bfa gene encoding a sequence-specific	
					DEFINITION selection	
					ACCESION L10036	
					REFERENCE 9142008	
					ARTICLE 1 (bases 1 to 4142)	
					SELECTOR Anabaena sp. strain PCC 7120 bfa gene cloned by in vivo transcriptional interference	
					JOURNAL J. Bacteriol. 175, 4025-4035 (1993)	
					MEDLINE 93308081	
					FEATURES Location/Qualifiers	
					SOURCE 1..4142	





OCCOYAGDRGRLOKAVTOORLIGTSIKIWEGDITQPGKKPPVELDRALLDVPSC GIGTILHNPDLNRQTATILLPLQQQLAKIAPIYKSGSTLYSTCTCINPAEAA Q1615..16837	CDSSource	Animale e Generica, Universite di Firenze, Via Romana 17, Florence, FI 50125, Italy
/note="ORF_ID:s1r0680" /codon_start=1 /product="hypothetical protein" /db_xref="PID:g1652036"		Location/Qualifiers 1..2123 /organism="Proteus mirabilis" /strain="AF 2924" <1..302 /note="orf1; similar to E.coli EnvM" /codon_start=3 /db_xref="PID:91053075" /transl_table=11 /product="glutathione transferase" /db_xref="PID:9103076" /transl_table=11 /translation=MKLWTPGSSLSRHTVLRTRGLDFDSIERIDLRRKTEGKDFL AINPRQYEVQLQDNGDLDLTGEGTAIVOLADEPDRNIAAPPALERHOIENFLA SEHKGYSPFESSDTPPESYLPVVKNLKSKFVINDVLSKQKVCVDIFTVADAYLF LSQWAPHAVALDDLSHQDYLARIAQRPNVHSALVTLIKE" complement="orf1..1738" /note="orf2" /codon_start=1 /db_xref="PID:91053077" /transl_table=11 /product="MGHEFEKGCLIPPEVSGVLCEDALPISDIJAPNLLELETTAGAT LHHVNDQCYAAAROLCOOPDKPKIVKHSRAGRFRDREMILYTADISWHVSPLYDGS EROPGVGDFTLDTSLMLVQMLSVELTKALEHYAAAYEVMLKTKENNEYELQVARD OMYHPTHEFCATQD"
Db 102039 tagctcgccgtgtggatattcccaactacacaactgtatggggttggcgaagccggtcctgg 102158	CDS	
Qy 383 TCGCTCATGAACCTAAAAAATTAATGCCAGAGGTGTTAGATTGTGCAACACATATT 442		
Db 102099 ttggccggcgtgtggatattcccaactacacaactgtatggggttggcgaagccggtcctgg 102098	CDS	
Qy 443 TAGTGTGGCAGANTTCAGTCAAATTAATGTTGAGCTTAAAGCGAGCTTAG 502		
Db 102159 aaatgtgttgtatctgtggcagaatgtggggcccaaataatttgcgttaatggta 102218	CDS	
Qy 503 AAGCAATATGTTAAATATTAGCATAGACTTAACTAGATAGCTGTATAATTTCGGTTATATGCRAA 562		
Db 102219 tttcgcgtggcccatccggacccatccggcatgtttgtatata 102278	CDS	
Qy 563 TTTCAGCTGGNCACAATCCGTCACATTAAGTCACAAGGTGNGGGTTCATACAAATTTC 622		
Db 102279 ttccaaatgtggaaagggtggcaccactaaaggcgtaaccgttacccaaacggaaatgtggga 102338	CDS	
Qy 623 TTAAAGAAAATCGAAAGCGCTGCAACCTTAAANACGTAACGGTGATAAGTAGAACTAGGA 682		
Db 102339 atacagcggttttcgtggccaaatgtggccatccatggcacaattatctcaag 102398	CDS	
Qy 683 AAACAGCGGCTTACTTAAAGTGACTTAACTCAAGTGGCTTACAGGTGAAATAATTTCATG 742		
Db 102399 tccat 102403	CDS	
Qy 743 TAGAT 747		
RESULT 3 PMU38482 2123 bp DNA BCT 06-SEP-1996	DEFINITION	Protein mirabilis glutathione transferase (gstB) gene, complete
ORGANISM Proteus mirabilis strain=AF 2924.	VERSION	91053074
SOURCE Proteobacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;	REFERENCE	U38482
ORGANISM Proteus mirabilis	AUTHORS	Mignogna, G., Allocati, N., Aceto, A., Piccolomini, R., Di Ilio,C., Barra, D. and Martini,F.
ORGANISM Bacterium	TITLE	The amino acid sequence of glutathione transferase from Proteus
ORGANISM Bacterium	JOURNAL	Journal of Bacteriology. 211 (3), 421-425 (1993)
REFERENCE 2 (bases 1 to 2123)	AUTHORS	Perito, B., Allocati, N., Casalone,E., Masulli,M., Dragani,B., Polinelli,M., Aceto, A. and Di Ilio,C.
REFERENCE 2 (bases 1 to 2123)	TITLE	Molecular cloning and overexpression of a glutathione transferase
REFERENCE 2 (bases 1 to 2123)	JOURNAL	Biochem. J. 318 (Pt 1), 157-162 (1996)
REFERENCE 3 (bases 1 to 2123)	AUTHORS	Perito, B. and Casalone, E.
REFERENCE 3 (bases 1 to 2123)	TITLE	Submitted (13-OCT-1995) Enrico Casalone, Dipartimento di Biologia
REFERENCE 3 (bases 1 to 2123)	JOURNAL	

AUTHORS	Fleischmann, R.D., Adams, M.D., White, O., Clayton, R.A., Kirchner, E.F., Kerlavage, A.R., Bult, C.J., Tomb, J.-F., Dougherty, B.A., Merrick, J.M., McKenney, K., Sutton, G., FitzHugh, W., Fields, C.A., Gocayne, J.D., Scott, J.D., Shirley, R., Liu, L.-I., Glodek, A., Kelley, J.M., Weidman, J.F., Phillips, C.A., Spriggs, T., Hedblom, E., Cotton, M.D., Utterback, T.R., Hanna, M.C., Nguyen, D.T., Sauder, D.M., Brandon, R.C., Fine, L.D., Fritchman, J.L., Fuhrmann, J.L., Geoghegan, N.S.M., Gnehm, C.L., McDonald, L.A., Small, K.V., Fraser, C.M., Smith, H.O. and Venter, J.C.	FEATURES	search capability are available at TIGR's World Wide Web site ( <a href="http://www.tigr.org/tdb/mdb/hmdb/hmdb.html">http://www.tigr.org/tdb/mdb/hmdb/hmdb.html</a> ). Location/Qualifiers 1..10111 /organism="Haemophilus influenzae" complement(1685..2581) /gene="HI1732" /note="similar to SP_003155 percent identity: 28.6; identified by sequence similarity; putative. E. Koonin: weak similarity to AIDA_ECOLI, BLAST score: 88" /codon_start=1 /product="adhésin (aida-I)" /db_xref="PID:91574389" /trans_table=11 /translation="NNKIFKVIVNVVOTWVVYSELTRAHTRKTSATVATLAVLS ATVOAINDASDEVKYSTEDDIEDSAATDDDNKNOALKAGDTTILKAGGNKKAKLDOG GKSIVTAFALARDLDVTKAVSPLTIGGNTPAAGGATPKVITSPTADGLAKGNGDT AVHNLGALSTLPDVTTNGASTSYVSFSPSDIEKTRAAIKDVLNAGWNKGAKYAGNN TENVNDLVAGIDVNVEETGDETDNTLYDVTAKENGKTEKTPKTSVIRDKNGKLTK QLKDANTGTAATNAEDAEAMA" complement(2937..4916) /gene="HI1733" /note="similar to SP_030850 percent identity: 50.8; identified by sequence similarity; putative. E. Koonin: homolog of RNB_ECOLI, BLAST score: 529" /codon_start=1 /product="exoribonuclease II (RNaseII)" /db_xref="PID:91574390" /trans_table=11 /translation="MFQDNPLLAQOLKQOIQHDSKKEQEVGVYKSTDKAYGFLEDKKTYF IAPPSSMKVNHGDKTKATIEKOGDKKEQAEPEALPEPMTRFIAKYRENDKKQQLVND WWVTLARHEQPYRPGAPPYEMLDOKTNRNLTALHFETMDSESTMMDALYEPIA HSPIINGPIGAQAKSVTKEEQDQHVVANQWVPPRDPDFEYFTINOLICRDBLAP QNSTQNGKLVVAIADPTAVIALDSQIEQAEAKCQTNYLPGFNIPMPLRESDELCS LJANEETPAPLCVYIEDLDTGNTAFLKPSYKQDQVYKNSDLYQADNAWQPE MPETACQIHWLHOFTKARIOWRKTHSLFPEKPDYAFVLAENGVQEKAKEYRLANO IVEEAMITIANICAQAFHEADKTGLENTISGEDKFKELAHNFELMANEQNQTELLA TWTSPKYSQDMVNERLKAYLAQKOPQYXPNQDYLARQGEARRRNVRDIAWLYC RYLSKDVASNAEEFEAEVDYMRAGLVRVOLENGSLFPAATLHNNEKEIQLNPDELA LYTIKGRTYKUGDMVKVQLTIEPQVATRSTIVGELQ" complement(4999..5886) /gene="HI1734" /note="similar to GB:M97219_1 percent identity: 75.3; identified by sequence similarity; putative. E. Koonin: homolog of FAB1_ECOLI, BLAST score: 978" /codon_start=1 /product="short chain alcohol dehydrogenase homolog (envM)" /db_xref="PID:91574591" /trans_table=11
JOURNAL	Science 269 (5223), 496-512 (1995)	SOURCE	
REFERENCE	2 (bases 1 to 10111)	CDS	
AUTHORS	L Tatusov, R., Mushegian, A.R., Bork, P., Brown, N.P., Hayes, W.S., Borodovsky, M., Rudd, K.E. and Koonin, E.V.		
TITLE	Metabolism and evolution of haemophilus influenzae deduced from a whole-genome comparison with escherichia coli		
JOURNAL	Curr. Biol. 6 (3), 279-291 (1996)	CDS	
EDITION	3 (bases 1 to 10111)		
REFERENCE	Fleischmann, R.D., Adams, M.D., White, O., Clayton, R.A., Kirchner, E.F., Kerlavage, A.R., Bult, C.J., Tomb, J.-F., Dougherty, B.A., Merrick, J.M., McKenney, K., Sutton, G., FitzHugh, W., Fields, C.A., Gocayne, J.D., Scott, J.D., Shirley, R., Liu, L.-I., Glodek, A., Kelley, J.M., Weidman, J.F., Phillips, C.A., Spriggs, T., Hedblom, E., Cotton, M.D., Utterback, T.R., Hanna, M.C., Nguyen, D.T., Sauder, D.M., Brandon, R.C., Fine, L.D., Fritchman, J.L., Fuhrmann, J.L., Geoghegan, N.S.M., Gnehm, C.L., McDonald, L.A., Small, K.V., Fraser, C.M., Smith, H.O. and Venter, J.C.		
AUTHORS	Fleischmann, R.D., Adams, M.D., White, O., Clayton, R.A., Kirchner, E.F., Kerlavage, A.R., Bult, C.J., Tomb, J.-F., Dougherty, B.A., Merrick, J.M., McKenney, K., Sutton, G., FitzHugh, W., Fields, C.A., Gocayne, J.D., Scott, J.D., Shirley, R., Liu, L.-I., Glodek, A., Kelley, J.M., Weidman, J.F., Phillips, C.A., Spriggs, T., Hedblom, E., Cotton, M.D., Utterback, T.R., Hanna, M.C., Nguyen, D.T., Sauder, D.M., Brandon, R.C., Fine, L.D., Fritchman, J.L., Fuhrmann, J.L., Geoghegan, N.S.M., Gnehm, C.L., McDonald, L.A., Small, K.V., Fraser, C.M., Smith, H.O. and Venter, J.C.		
TITLE	Submitted (27-JUL-1995) The Institute for Genomic Research, 9712		
JOURNAL	Direct Submission		
REFERENCE	4 (bases 1 to 10111)		
AUTHORS	White, O., Clayton, R.A., Kerlavage, A.R. and Fleischmann, R.D.		
TITLE	Submitted (27-SEP-1996) The Institute for Genomic Research, 9712		
JOURNAL	Medical Center Dr., Rockville, Maryland 20850, USA		
REMARK	The H. influenzae sequence has been updated by R. Fleischmann. New database matches have been assigned, product names have been improved, and a number of frame shifts have been corrected. We gratefully acknowledge the work of Tatusov et. al. We have incorporated their annotation into the /notes field of the corresponding H. influenzae genes.		
COMMENT	Address all correspondence to: J. Craig Venter The Institute for Genomic Research 9712 Medical Center Dr., Rockville, MD, 20830, USA		
	Hamilton O. Smith is with the Johns Hopkins University School of Medicine, Baltimore, MD, 21205.		
	Joseph M. Merrick is with the State University of New York, Department of Microbiology, Buffalo, NY, 14214.		
	Chris Fields' current address is the National Center for Genome Resources, Santa Fe, NM, 87505.		
	All other authors are with The Institute for Genomic Research, 9712 Medical Center Dr., Rockville, MD, 20850.		
	Predicted open reading frames were determined using GenMark software, kindly supplied by Mark Borodovsky, Georgia Institute of Technology, Atlanta, GA, 30332.		
	e-mail: mark@amber.gattech.edu		
	Submission and annotation: Owen White		
	e-mail: owhite@tigr.org		
	Curation of the genome is ongoing and comments to the authors are appreciated. Biological role information, putative identifications, sequence alignments, on-line name and sequence		











Headed by:  
 Name: Takashi Horiochi  
 Address: National Institute of Basic Biology, Okazaki, 444 Japan  
 E-mail: kishorin@nibb.ac.jp  
 Information operator:  
 Name: Hirotada Mori  
 Address: NARA Institute of Science and Technology,  
 E-mail: hmori@tgc.aist-nara.ac.jp  
 URL: http://ds3.aist-nara.ac.jp/  
 The Japan E. coli genome database  
 http://ds3.aist-nara.ac.jp/  
 Location/Qualifiers

**FEATURES**

source	1..17297	/organism="Escherichia coli"
		/note="Nucleotide position 134 053-1357859 from the initiation site of Thra (0 min.); This clone is from Kohara lambda miniset library"
		/strain="K1"
		/clone="Kohara clone #255"
		/map="28.8 min"
	<1..771	
		/gene="acna"
		/note="ORF_ID:o254#5#1; similar to [PIR Accession Number S22375]
		/codon_start=1
		/product=Aconitase hydratase (EC 4.2.1.3)"
		/db_xref="PID:91742090"
		/transl_table=11
		/translation="DSTYIFLSPPFDEMOATPAPVEDIHGARIALMLGDSTYTDHISP MTRHUPDSVYISIVDAAMRKOETPLSYAKTGYSGSSRDVAKGRLGRVVIA ESTERHRSHNLIGNLPLPEQFQSTTRKGLIGLSEKIDGDLQNLQFGATPVTLIR ADGSQEPVCPRCRIDATELYYQNDGLHRYVNMLK"
		/complement:835..1425
		/gene="ribo"
		/note="ORF_ID:o255#2; similar to [PIR Accession Number A40654]"
		/codon_start=1
		/product="GTP cyclohydrolase II (EC 3.5.4.25)"
		/db_xref="PID:91742091"
		/transl_table=11
		/translation="MOLKRVAEAKLPTPWGDFLMYFEELATGHDHVAYGDISHT PVLYRHSECLGDAFLSLRGFOLEALTAQTBEGRILLYHRQEENIGLNNKI RAYALOGYDVEANHOLGAFAADEDEFTCADEMKLIGNEVRLLTNPKVBEILTE 1595..2359
		/gene="pgpB"
		/note="ORF_ID:o255#3; similar to [SwissProt Accession Number P18201]"
		/codon_start=1
		/product="phosphatidylglycerophosphatase B (EC 3.1.3.27)." /db_xref="PID:91742092"
		/transl_table=11
		/translation="MRSIARATTAVGALLIYMPVAVISGWRWGEOSWILKAFAFWV TEVTPQMGVIHLIFGWFLNCLPRRIKAFLVLAAILYQOGVKSWIKYVOE PREFVIAWVLAQVHWRPRTLTIAILWVATCYSMSRLGMHWPDLVVATLI 2508..2616
		/note="ORF_ID:o255#4; similar to [SwissProt Accession Number P44129]"
		/codon_start=1
		/db_xref="PID:91742093"
		/transl_table=11
		/translation="MKTLYFLLVLAIFVISTVLGAONQQVTNYLLAQEYRISTL LAVLFAAGFAIGMILICGFLWLRVRSALARERIKRLENQSPATDVAVPHSSAAKE 2823..3902
		/gene="ygiM"
		/note="ORF_ID:o255#5; similar to [SwissProt Accession Number P45576]"

Note: remainder of annotations omitted.



Members: (1995.4 - 1996.3)  
 Aiba H., Baba T., Fujita K., Hayashi K., Honjo A., Horiochi T., Ikemoto K., Inada T., Isono K., Isono S., Kimura S., Kitagawa M., Kitamoto K., Kim S., Masuda S., Miki T., Mizobuchi K., Mori H., Motomura K., Nakamura Y., Nashimoto H., Nishio Y., Oshima T., Saito N., Sampei G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y., and Yan, M.

Headed by:  
 Name: Tatsushi Horiuchi  
 Address: National Institute of Basic Biology, Okazaki, 444, Japan  
 E-mail: kishorien@nibb.ac.jp  
 Information operator: hirota.mori@nara.ac.jp

Address: NARA Institute of Science and Technology,  
 Ikoma, 630-01, Japan  
 URL: http://bsw3.aist-nara.ac.jp

Source  
 /organism="Escherichia coli"  
 /strain="K12"

Location/Qualifiers  
 1..17769  
 /codon\_start=1  
 /note="Nucleotide position 1347168-1364936 from the initiation site of Thra (0 min.); This clone is from Kohara lambda miniset library"  
 /clone="Kohara clone #256"  
 /clone\_id="Kohara lambda miniset library"  
 /map="29.0 min"  
 complement(421..2406)  
 /gene="MTCYB02B10\_18c"  
 /note="ORF\_ID:0255#12; similar to [SwissProt Accession Number Q11024]"  
 /codon\_start=1  
 /db\_xref="PID:91742108"  
 /transl\_table=11

Translation="MKTVERSTLTYNLFGLSHNPYWRLLTESSDVLRFSTTETTEDRTLQLSAEAAIRKEVTVLQVLSLIVLGRKINREWAGNASAWHDTPAVARDLSSHLGSLPAEKFQVSEAHSAVILUDSGRNQIREFRLCEDYTGKHEHDVIGOSVYKLFMSRREAAASPRNNRVEFRRSENAVEYLWISTCKGQRLEPFRKFWHSGSCKKNFPLICFKVNDAYGHIFHGQDQLRDLTSLSLUSCLENDDQGLARPSGDEFLVLAISNTSQSALAEAMA SRILUTBLRPERIGLIEVYVSCSYGTLSPSEGSDSTAIIRHADTMTTAKEGRGQFQCVTFPMNQRTEYLWDLNQHDLVTHQDKITWGRGLVWDRGQINLYAVNTSAROLADOTIPTAOKVOLNFXCPIDVLETSCLIENDELASVQOESQLGQWHLDGFTGYSLSLSQALRHNKEADGVATEMLDGLVREDLTALDFTLDSASTEDMDALPAKALPDDKLOLIVATAOPTAWIAKSMLDGAATKPAENTNLGNMPRELSDDLCRLANEVPLACRMLSAUGTIEDNIEFFAATESKAKLYDQSVWLENTGDPPESEAIAEQVLLAQICQRGEWRHNALEADGPDRPYFLUGKEPKVLDVAEPRLANIVEEA MIANANIAQVLPRLKDFGEGKLVNGWNGDPAANDALALKTHGDAEWFLDGFCKLRLDAQFGLDSRIRFQSFAEISTEPGPHGLBLAYAWTSPTRKYGMINHRLKAVIKGETATRQEDETOVMAERURUNRMARDVGWLYAFKUKAGTURFRAAEIVDLSRGMVRYAVDNGAIAFAPLARVDELVCSENENGVOIKEVIVKTDVIDTVAEVRMETRSIATRPVA"  
 complement(5914..6699)  
 /gene="fabI, envM"

CDS  
 /note="ORF\_ID:0255#16; similar to [SwissProt Accession Number P29132]; Author-given protein sequence is in conflict with the conceptual translation"  
 /codon\_start=1  
 /product="Enoyl-[acyl]-carrier-protein] reductase (NADH) (EC 1.3.1.9) (NADH-dependent enoyl-ACP reductase)."  
 /db\_xref="PID:91742110"  
 /transl\_table=11  
 /translation="GFLSGSKRILYVGASKLUSIAYGIAQMHRGAEALLAFTYQNDKLGRVEFAQGSDIVLQCDVAEDASIDMFAELGKWPFDGFVHSIGFAPGDQDLDGVNAYTRREGKTAHDISISSFVNAMAKRSMPLNSALITLSYGAERALPNVNMGAKASUEANRYMANAMGPEPSVYRNASAGSPIRTLAASGAGISGEVVHVDGGFIAAMNELELK"  
 complement(7070..7423)  
 /gene="ycld"  
 /note="ORF\_ID:0255#17; similar to [SwissProt Accession Number P45736]"  
 /codon\_start=1  
 /db\_xref="PID:91742111"  
 /transl\_table=11  
 /translation="MIDKTKSNARDLRRNNTLQERKLWYLRSRBRESDFEKERROHPVSYLDFACCSARYVYELDGGQHDLAVAVYDSRTSRTSWLESSGWTVRWNNEIDCNEETVLENLQELNRSPSP"  
 complement(7491..8297)  
 /gene="spf"  
 /note="ORF\_ID:0255#18; similar to [SwissProt Accession Number P46637]"  
 /codon\_start=1  
 /product="Peptide transport system ATP-binding protein SapF"  
 /db\_xref="PID:91742112"  
 /transl\_table=11  
 /translation="MIETLIEVRNLSKTRFYRTGWERROTVEAYKPLSLSTLREGOTLA  
 LIGENSGGSTLAMLAGIIEPTGEGELDDDHPLHEFGDYSERSRSTRMFQDPSLSPRQISQDPLRNTDEPEQRKQIPLDHHVQYPPMAQGQKQRLGLARALIIPRVTIADEALSLDMEMRSQIUNLMLQEQGISTYVTCIGHAMKHISDQVLYMHOCEVVERGNGSTAVYLASPHELTKRLTAGHFGSALTADAWRKR"  
 complement(8229..9491)  
 /gene="spd"  
 /note="ORF\_ID:0255#19; similar to [SwissProt Accession Number P36635]"  
 /codon\_start=1  
 /product="Peptide transport system ATP-binding protein SapD (fragment)"  
 /db\_xref="PID:91742113"  
 /transl\_table=11  
 /translation="MPILLDIRNLNTIEFKTGDENVKADEVRSMLTGEFLRGVLVGSGS  
 GKSLIAKACGVNNDNWRVTAIDMRNFDLDRLLSAREERKLVCHNMSNIFQEFQSCLL  
 DPSERVGRGOMNIPAWTKGRWQRFGRKRAEJFHRVJKDHKUPMRSFYELET  
 EGECOKYMATIALANOPRLIADEPNTNSEMEPTQOQIFLILTRUNQNSNTILLSHD  
 LQMUSQWAKILINVLCQGVETAPSKELTMPPHYPTQOLIARATPDFGAMPKRSRL  
 TLPEAIPLEQLPQGCRUQPRCPQARECIVTPPLTGARNHLYACHFPLNMEKE"  
 complement(9291..10181)  
 /note="ORF\_ID:0255#20; similar to [SwissProt Accession Number P36669]"  
 /codon\_start=1  
 /product="Peptide transport system permease protein SapC."  
 /db\_xref="PID:91742114"  
 /transl\_table=11  
 /translation="MPYDVSYVSEKRPPGTLLRTAWKFKYSDASAMVGLYGGAGLAVLCLFGWMFAPYDQDOFLGYQOLQPPWSFLGTDIGRDYLSRLGSAAPTUVGGA  
 FVYVLLAATCGLVLTGTFGATHGURSAVNHILDTLAIPSLLIAITVYAFAGPSL  
 AMFRAWLALLPRAVNSIXYMHDELEKEVVIARLDGASTNLILWFAMPNTAGLV  
 ETIRALSMALDIAALGFDLGQOLPSPGAMELGDAELLELIYAPWTTMLPGAIMIS  
 VLLVNLIGVYRRAIAGE"  
 complement(10168..11133)  
 /note="ORF\_ID:0257#1; similar to [SwissProt Accession Number P36668]"  
 /codon\_start=1  
 /product="Peptide transport system permease protein SapB."  
 /db\_xref="PID:91742115"  
 /transl\_table=11  
 /translation="MIFTLRILLIVTFLFTVGFSLSYFTPHAPIQGASLWNW"

CDS  
 /note="ORF\_ID:0255#11; similar to [SwissProt Accession Number P36669]"  
 /codon\_start=1  
 /product="Exoribonuclease II (EC 3.1.13.1) (Ribonuclease III) (RNase III)"  
 /db\_xref="PID:91742109"  
 /transl\_table=11  
 /translation="MFQDNPLLAOLKQOLHQSQTPTAEGVYKATEKGFGFLEVDAQKSYFIPPPOMKKYMHGDLTIAVTHSEKRESEAPELYVPEFLITVKGVQGANDRAIVPDHPLKDAIPCRARGNHEKGDWAVAEKRRIPIKQDGSFYAETQITFDDHFPVWWVTLARHNLKEADGVATEMLDGLVREDLTALDFTLDSASTEDMDALPAKALPDDKLOLIVATAOPTAWIAKSMLDGAATKPAENTNLGNMPRELSDDLCRLANEVPLACRMLSAUGTIEDNIEFFAATESKAKLYDQSVWLENTGDPPESEAIAEQVLLAQICQRGEWRHNALEADGPDRPYFLUGKEPKVLDVAEPRLANIVEEA MIANANIAQVLPRLKDFGEGKLVNGWNGDPAANDALALKTHGDAEWFLDGFCKLRLDAQFGLDSRIRFQSFAEISTEPGPHGLBLAYAWTSPTRKYGMINHRLKAVIKGETATRQEDETOVMAERURUNRMARDVGWLYAFKUKAGTURFRAAEIVDLSRGMVRYAVDNGAIAFAPLARVDELVCSENENGVOIKEVIVKTDVIDTVAEVRMETRSIATRPVA"  
 complement(5914..6699)  
 /gene="fabI, envM"







vector=Bluescript host=Golden Hamster Rsite1-EcoRI Rsite2-XbaI Total mRNA was isolated from adult worms using the acid guanidine thiocyanate method (Chomczynski and Sacchi, 1987). Poly (A)+ RNA was purified by affinity chromatography on oligo dT column (Aviv and Leder, 1972). The cDNA expression library was constructed using the Lambda Zap cDNA cloning system (Stratagene-[Uni-Zap XR Gigapack II cloning kit]). The library was amplified (end titer=0.5 x 10<sup>8</sup> pfu/ml) and its efficiency tested for using known *S.mansoni* probes (chorion and troponin genes). The size range of inserts for the library was 0.5kb - 2.5 kb, the cDNA being directionally cloned between the EcoRI and the XbaI sites of the Bluescript phagemid. Bluescript phagemid was excised by mass in vivo excision of the library and inserted into E.coli XL1 Blue. The plasmid was purified using the Qiagen Qiawell 8 system and sequenced with Taq dye-terminator chemistry (SK-73 and T7 primers) for ABI 373A. The library was prepared by Hanan Abdel Hameed, Ain Shams Univ, Cairo and provided by Mohamed Saber, Theodore Bilharz Research Institute, Cairo.

**ORGANISM** Schistosoma mansoni  
**Eukaryota; mitochrondrial eukaryotes; Metazoa; Platyhelminthes; Trematoda; Digenea; Strigeida; Schistosomatidae;**

**AUTHORS** I. Ridgers, I. L.

**COMMENT** Unpublished (1995)

**REFERENCE** Ridgers, I.L.  
**AUTHORS** I. Ridgers  
**COMMENT** Experimental Taxonomy Unit, Zoology  
 The Natural History Museum  
 Cromwell Road, South Kensington, London, SW7 5BD, U.K.  
 Tel: 01719389297  
 Fax: 01719388754

**FEATURES** source  
**REFERENCE** I.Ridgers@nhm.ac.uk  
**AUTHORS** Schistosoma mansoni cDNA clone SNHAD016014T3.  
**COMMENT** Location/Qualifiers  
 1..337  
*/organism="Schistosoma mansoni"*  
*/strain="Egyptian"*  
<1...>337

**BASE COUNT** 85 a 58 c 56 g 121 t 17 others

**RESULT** 2  
**DEFINITION** AA039422 439 bp mRNA EST 30-AUG-1996

**ACCESSION** AA039422  
**NID** 91515639

**KEYWORDS** EST  
**ORGANISM** Homo sapiens

**REFERENCE** Holman, M., Hulman, T., Clark, N., Dubuque, T., Elliston, K., Hawkins, M., Parsons, J., Rirk, J., Treviskis, E., Waterston, R., Williamson, A., Wohldmann, P., and Wilson, R.  
**AUTHORS** The WashU-Merck EST Project  
**COMMENT** Unpublished (1995)

**RESULT** 2  
**DEFINITION** AA039422 439 bp mRNA EST 30-AUG-1996

**ACCESSION** AA039422  
**NID** 91515639

**KEYWORDS** EST  
**ORGANISM** Homo sapiens

**REFERENCE** Holman, M., Hulman, T., Clark, N., Dubuque, T., Elliston, K., Hawkins, M., Parsons, J., Rirk, J., Treviskis, E., Waterston, R., Williamson, A., Wohldmann, P., and Wilson, R.  
**AUTHORS** The WashU-Merck EST Project  
**COMMENT** Unpublished (1995)

**FEATURES** source  
**REFERENCE** Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: est@watson.wustl.edu  
 This clone is available royalty-free through LINL ; contact the IMAGE Consortium (info@image.lnl.gov) for further information.  
 Seq. Primer: -28M13 rev2 from Amerham  
 High quality sequence stop: 246.  
**LOCATION/QUALIFIERS**  
 1..439  
*/organism="Homo sapiens"*  
*/note="Organ: uterus; Vector: pT7M3-PAC; Site\_1: Not I; Site\_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dt) primer [5'-AATCAGAATTTCGCGCCATTTTTTTTTTTTT 3'],"*  
*double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7M3 vector. Library went through one round of normalization. Library constructed by M. Fatima Bonaldo."*  
*/clone="J85191"*  
*/clone\_id="Soares pregnant uterus NBHPU"*  
*/sex="female"*  
*/dev\_stage="adult"*  
*/lab\_host="DH10B"*  
<1...>439

**FEATURES** mRNA  
**BASE COUNT** 125 a 81 c 79 g 145 t 9 others  
**ORIGIN**  
**QUERY MATCH** 3.0%  
**BEST LOCAL SIMILARITY** 71.7%  
**MATCHES** 38; Conservative 0; Mismatches 15; Indels 0; Gaps 0;  
**DEFINITION** J9617F Fetal heart, Lambda ZAP Express Homo sapiens cDNA clone  
**ACCESSION** N86668  
**NID** 91439870  
**KEYWORDS** EST  
**ORGANISM** Homo sapiens  
**REFERENCE** 1 (bases 1 to 179)  
**AUTHORS** Liew CC  
**TITLE** CDNAS from fetal heart (1996)  
**JOURNAL** Unpublished (1996)  
**COMMENT** Contact: Liew CC  
 Molecular Cardiology  
 University of Toronto  
 Banting Institute, 100 College St., Toronto, Ontario, M5G1L5  
**FEATURES** source  
**REFERENCE** I. (bases 1 to 439)  
**AUTHORS** Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M., Holman, M., Hulman, T., Lennon, G., Marra, M., Parsons, J., Rirk, J., Treviskis, E., Waterston, R., Williamson, A., Wohldmann, P., and Wilson, R.  
**TITLE** The WashU-Merck EST Project  
**JOURNAL** Unpublished (1995)

**FEATURES** source  
**REFERENCE** XhoI; mRNA was purified from human fetal hearts (8-10 weeks). cDNA was synthesized using a XhoI-Oligo dT adapter primer. EcoRI adaptors were ligated, followed by predigested lambda ZAP Express.  
*/clone="J9617"*  
**FEATURES** source  
**REFERENCE** /organism="Homo sapiens"  
*/note="Vector: Lambda ZAP Express; Site\_1: EcoRI Site\_2: XhoI; mRNA was purified from human fetal hearts (8-10 weeks). cDNA was synthesized using a XhoI-Oligo dT adapter primer. EcoRI adaptors were ligated, followed by predigested lambda ZAP Express."*  
*/clone="J9617"*

Db	101	ggtgtatgtat	112					
Qy	265	GGTGTATCAT	276					
RESULT	5	AA138421	424 bp	mRNA	EST	02-DEC-1996		
LOCUS		mc88b09 r1	Stratagene mouse melanoma (#337312)	Mus musculus	EST			
DEFINITION		clone 58575 5'	similar to SW:DLDH_MYCGE P47513 LIPOAMIDE DEHYDROGENASE COMPONENT					
ACCESSION		AA138421						
NID		9170622						
KEYWORDS								
SOURCE		house mouse.						
ORGANISM		Mus musculus						
KEYWORDS		Eukaryote; mitochondrial eukaryotes; Metazoa; Chordata; Vertebrata; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus						
REFERENCE		1 (bases 1 to 424)						
AUTHORS		Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T., Geiselt, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M., Scheibenbogen, K., Steptoe, M., Tan, F., Underwood, K., Moore, B., Theisinger, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and Waterston, R.						
TITLE		The WashU-HMMI Mouse EST Project						
JOURNAL		Unpublished (1996)						
COMMENT		Contact: Marra M/Mouse EST Project WashU-HMMI Mouse EST Project Washington University School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108 tel: 314 286 1800 Fax: 314 86 1810 Email: mouseest@watson.wustl.edu						
FEATURES		This clone is available royalty-free through LNLN ; contact the IMAGE Consortium (info@image.lnl.gov) for further information.						
BASE COUNT		MG1: 360433	Seq primer: primer name ambiguous					
ORIGIN		High quality sequence stop: 416.						
source		Location/Qualifiers						
		1..424						
		/organism="Mus musculus"						
		/note="Vector: pBluescript SK-; site_1: EcoRI; site_2: XbaI; Cloned unidirectionally. Primer: Oligo dT. From M cells, highly metastatic derivative of the K-1735 (mouse) melanoma. Average insert size: 1.0 kb; Uni-ZAP X vector; -5', adaptor sequence: 5' GAATTCTGGCACGAG 3'; -3' adaptor sequence: 5' CTCGAATTCTGGCACGAG 3'; "clones"="585785"						
		/clone.lib#="Stratagene mouse melanoma (#937312)"						
		/dev_stage="M2" cells"						
		/lab.host="SOLR (kanamycin resistant)"						
		<1..>424						
		175 a 48 C 74 g 127 t						
RESULT	6	W57951	435 bp	mRNA	EST	15-OCT-1996		
LOCUS		z221e07.r1	Soares fetal heart	NbHH119W	Homo sapiens	cdNA clone		
DEFINITION		341316 5'.						
ACCESSION		W57951						
NID		91364733						
KEYWORDS		EST.						
SOURCE		human.						

ORGANISM	Homo sapiens	Wilson, R.
Vertebrate; Mammalia; Chordata;	The WashU-Merck EST Project	
Eukaryota; Eutheria; Primates; Catarrhini; Hominidae; Homo.	Unpublished (1995)	
REFERENCE	1 (bases 1 to 435)	COMMENT
AUTHORS	Hillier,L., Clark,N., Dubuge,T., Elliston,K., Hawkins,M., Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M., Parsons,J., Riffkin,L., Rohlfing,T., Soares,M., Tan,F., Trevaskis,E., Waterston,R., Williamson,A., Wohldmann,P. and Wilson,R.	CONTACT: Wilson RK WashU-Merck EST Project Washington University School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108 Tel: 314 286 1800 Fax: 314 286 1810 Email: estwatson@wustl.edu
TITLE	The WashU-Merck EST Project	FEATURES
JOURNAL	Unpublished (1995)	SOURCE
COMMENT	This clone is available royalty-free through LNLN ; contact the IMAGE Consortium (info@image.lnl.gov) for further information.	organism="Homo sapiens" note="Orga: heart; Vector: pT7T3D (Pharmacia) with a modified polylinker; Site:1: Not I; Site:2: Eco RI; 1st strand cDNA was primed with a Not I - Oligo(dT) primer [5'-TGTTACCAATCGAACTGGGAGCGCATTTCCTTTTTTTT-3'], double-stranded cDNA was size selected, ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of a modified pT7T3 vector (Pharmacia). Library went through one round of normalization to a Cot = 5. Library constructed by M.Fatima Bonaldo. This library was constructed from the same fetus as the fetal lung library. Soares fetal lung NbHH19W."
FEATURES	Insert Length: 935 Std Error: 0.00 Seq Primer: mob REGA-ET High quality sequence stop: 338.	/clone="343004" /clone.lib="Soares fetal heart NbHH19W" /sex="Unknown" /dev_stage="19 weeks" /lab_host="DH10B (ampicillin resistant)"
JOURNAL	Contact: Wilson RK WashU-Merck EST Project Washington University School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108 Tel: 314 286 1800 Fax: 314 286 1810 Email: estwatson@wustl.edu	BASE COUNT
COMMENT	This clone is available royalty-free through LNLN ; contact the IMAGE Consortium (info@image.lnl.gov) for further information.	mRNA ORIGIN
AUTHORS	1. 435 /organism="Homo sapiens" /note="Organ: heart; Vector: pT7T3D (Pharmacia) with a modified polylinker; Site:1: Not I; Site:2: Eco RI; 1st strand cDNA was primed with a Not I - Oligo(dT) primer [5'-TGTTACCAATCGAACTGGGAGCGCATTTCCTTTTTT-3'], double-stranded cDNA was size selected, ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of a modified pT7T3 vector (Pharmacia). Library went through one round of normalization to a Cot = 5. Library constructed by M.Fatima Bonaldo. This library was constructed from the same fetus as the fetal lung library. Soares fetal lung NbHH19W" /clone="343116" /clone.lib="Soares fetal heart NbHH19W" /sex="Unknown" /dev_stage="19 weeks" /lab_host="DH10B (ampicillin resistant)"	Query Match Best Local Similarity 64.9%; Matches 48; Conservative 0; Indels 0; Gaps 0;
FEATURES	Insert Length: 935 Std Error: 0.00 Seq Primer: mob REGA-ET High quality sequence stop: 338.	Db Qy
JOURNAL	Unpublished (1995)	Base Count
COMMENT	163 a 84 c 67 g 121 t mRNA ORIGIN	163 a 84 c 67 g 121 t mRNA ORIGIN
AUTHORS	2. 9% Score 22; DB 170; Length 435; 1st Local Similarity 80.0%; Pred. No. 5.54e-04; Matches 36; Conservative 0; Mismatches 8; Indels 1; Gaps 1;	Query Match Best Local Similarity 64.9%; Matches 48; Conservative 0; Indels 0; Gaps 0;
FEATURES	b 301 agtgc当地atgttatgtcg-tttcaatcaaataattasagaatac 344 Y 589 ATGTGCAAAAGGTGTTGGGGGTCAATCAATTTAGAACANT 633	Db Qy
JOURNAL	Unpublished (1995)	RESULT
COMMENT	EST 15-OCT-1996 LOCUS w67124 516 bp mRNA DEFINITION 2d39a1.r1 Soares fetal heart NbHH19W Homo sapiens CDNA clone 291949 5' similar to SW:UBPX HUMAN P40818 PROBABLE UBIQUITIN CARBOXYL-TERMINAL HYDROLASE ; ACCESSION W02311 NID 91274290 KEYWORDS EST. ORGANISM Homo sapiens REFERENCE Hillier,L., Clark,N., Dubuge,T., Elliston,K., Hawkins,M., Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M., Parsons,J., Riffkin,L., Rohlfing,T., Soares,M., Tan,F., Trevaskis,E., Waterston,R., Williamson,A., Wohldmann,P. and Wilson,R. AUTHORS The WashU-Merck EST Project Unpublished (1995)	LOCUS W02311 604 bp mRNA DEFINITION 2a08dd7.r1 Soares melanocyte 2NBHM Homo sapiens CDNA clone 291949 5' similar to SW:UBPX HUMAN P40818 PROBABLE UBIQUITIN CARBOXYL-TERMINAL HYDROLASE ; ACCESSION W02311 NID 91274290 KEYWORDS EST. ORGANISM Homo sapiens REFERENCE Hillier,L., Clark,N., Dubuge,T., Elliston,K., Hawkins,M., Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M., Parsons,J., Riffkin,L., Rohlfing,T., Soares,M., Tan,F., Trevaskis,E., Waterston,R., Williamson,A., Wohldmann,P. and Wilson,R. AUTHORS The WashU-Merck EST Project Unpublished (1995)
FEATURES	7 W67124 516 bp mRNA OCUS 91376014 DEFINITION EST KEYWORDS human. ORGANISM Homo sapiens Eukaryota; mitochondrial eukaryotes; Metazoa; Chordata; Vertebrate; Eutheria; Primates; Catarrhini; Hominidae; Homo. REFERENCE Hillier,L., Clark,N., Dubuge,T., Elliston,K., Hawkins,M., Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M., Parsons,J., Riffkin,L., Rohlfing,T., Soares,M., Tan,F., Trevaskis,E., Waterston,R., Williamson,A., Wohldmann,P. and Wilson,R. AUTHORS The WashU-Merck EST Project Unpublished (1995)	RESULT
JOURNAL	Unpublished (1995)	ID SOURCE



Best Local Similarity 70.6%: Matches 36; Conservative 0; Mismatches 15; Indels 0; Gaps 0;	TITLE JOURNAL COMMENT	Takaichi,A., Takeda,S., Watanabe,T., Takahashi,E.-I., Hirai,Y., Maekawa,H., Shin,S. and Nakamura,Y. Unpublished (101) Unpublished (1995) Submitted (30-May-1995) to DDBJ by: Tsutomu Fujiwara Otsuka GEN Research Institute Otsuka Pharmaceutical Co.,Ltd 463-10 Kagasuno Kawachi-cho Tokushima, Tokushima Japan Phone: 0886-65-2888 Fax : 0886-37-1035.
RESULT 11 LOCUS AA092855 210 bp mRNA DEFINITION 119883-seq.F Fetal heart, Lambda ZAP Express Homo sapiens cDNA 5'. ACCESSION AA092855 NID 91637868 KEYWORDS EST. SOURCE human. ORGANISM Eukaryota; mitochondrial eukaryotes; Metazoa; Chordata; Vertebrata; Eutheria; Primates; Catarrhini; Hominidae; Homo. COMMENT REFERENCE 1 (bases 1 to 210) AUTHORS Liw.C.C. TITLE cDNAs from fetal heart (1996) UNPUBLISHED Unpublished (1996)	FEATURES source 1..214 /organism="Homo sapiens" /clone_id="Clontech human fetal brain polyA+ mRNA (#6535)"	BASE COUNT 35 a 32 c 54 g 54 t 39 others ORIGIN
RESULT 13 LOCUS HUM238E03B 233 bp mRNA DEFINITION Human aorta cDNA 5'-end GEN-238E03 . ACCESSION D62108 NID 9365884 KEYWORDS EST (expressed sequence tag); Human aorta; similar to none (May 29, 1995). SOURCE Homo sapiens (library: Clontech human aorta polyA+ mRNA (#6572)) ORGANISM Homo sapiens Eukaryota; mitochondrial eukaryotes; Metazoa; Chordata; Vertebrata; Eutheria; Primates; Catarrhini; Hominidae; Homo. COMMENT REFERENCE 1 (bases 1 to 233) AUTHORS Fujiwara,T., Hirano,H., Katagiri,T., Kawai,A., Kuga,Y., Nagata,M., Okuno,S., Ozaki,K., Shimizu,F., Shimada,Y., Shinomiya,H., Takaichi,A., Takeda,S., Nakamura,Y., Maekawa,H., Shin,S. and Nakamura,T., Unpublished (303) Submitted (30-May-1995) to DDBJ by: Tsutomu Fujiwara Otsuka GEN Research Institute Otsuka Pharmaceutical Co.,Ltd 463-10 Kagasuno Kawachi-cho Tokushima, Tokushima Japan Phone: 0886-65-2888 Fax : 0886-37-1035.	FEATURES source 1..233 /organism="Homo sapiens" /clone_id="Clontech human aorta polyA+ mRNA (#6572)"	BASE COUNT 91 a 34 c 32 g 69 t 7 others ORIGIN
RESULT 12 LOCUS HUM003B01A 214 bp mRNA DEFINITION Human fetal brain cDNA 3'-end GEN-003B01. ACCESSION D5961 NID 9960367 KEYWORDS EST (expressed sequence tag); Human fetal brain; similar to none (May 29, 1995). SOURCE Homo sapiens (library: Clontech human fetal brain polyA+ mRNA (#6535)) cDNA to mRNA. ORGANISM Eukaryota; mitochondrial eukaryotes; Metazoa; Chordata; Vertebrata; Eutheria; Primates; Catarrhini; Hominidae; Homo. COMMENT REFERENCE 1 (bases 1 to 214) AUTHORS Fujiwara,T., Hirano,H., Katagiri,T., Kawai,A., Kuga,Y., Nagata,M., Okuno,S., Ozaki,K., Shimizu,F., Shimada,Y., Shinomiya,H.,	FEATURES source 1..233 /organism="Homo sapiens" /clone_id="Clontech human aorta polyA+ mRNA (#6572)"	BASE COUNT 91 a 34 c 32 g 69 t 7 others ORIGIN

QY 383 TGGCTCATGAGCTAAATAATGCCAGAGGTGA 421

RESULT 14 mRNA EST 22-MAY-1995

LOCUS R44860 284 bp mRNA EST

DEFINITION y925al2.s1 Homo sapiens cDNA clone 33372 3.

NID 934233

KEYWORDS human clone=33372 library=Soares infant brain INIB vector=Lafnhd BA

host=DH10B (ampicillin resistant) primer=Promega -2In3 Rsite=Not I Rsite=Hind III Whole brain from a 73 days post natal female. 1st

strand cDNA was primed with a Not I - oligo(dT) primer [5', ACTGGAGAATTCGGCCGGCGACGAGTTTTTTTTTTT 3']; double-stranded

cDNA was ligated to Hind III adaptors (Pharmacia), digested with

Not I and directionally cloned into the Not I and Hind III sites of

the Lafnhd BA vector. Library went through one round of

normalization. Library constructed by Bento Soares and M.Fatima

Bonaldo.

Homo sapiens

Eukaryota; Metazoa; Chordata; Vertebrata; Gnathostomata; Mammalia;

Buteraria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 294)

AUTHORS Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,

Sarcopterygii; Chondriata; Vertebrata; Gnathostomata; Osteichthyes;

Eutheria; Archonta; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 284)

Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,

Holman,M., Hiltzman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,

Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F., Trevaskis,E., Waterston,R., Williamson,A., Wohldmann,P. and

Wilson,R.

The WashU-Merck EST Project

Unpublished (1995)

TITLE The WashU-Merck EST Project

JOURNAL GDB: G00-405-719

COMMENT Contact: Wilson RK

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: est@wustl.edu

High quality sequence stops: 271

Source: IMAGE Consortium, LBL

This clone is available royalty-free through LBL ; contact the

IMAGE Consortium (info@image.lnl.gov) for further information.

FEATURES Location/Qualifiers

SOURCE /organism="Homo sapiens"

/clone="116989"

BASE COUNT 97 a 62 c 49 g 85 t 1 others

ORIGIN ORIGIN

Query Match 2.7%; Score 21; DB 10; Length 294;

Best Local Similarity 88.9%; Pred. No. 7.94e-03; Mismatches 0; Indels 0; Gaps 0;

Matches 24; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Db 256 taatattaaatgtacttcaactgttaaga 282

QY 99 TAAATGATATTACITTAACCTAACAA 125

Search completed: Fri Dec 5 08:08:05 1997

Job time : 502 secs.

FEATURES SOURCE

ORIGIN /organism="Homo sapiens"

/clone="33372"

BASE COUNT 95 a 38 c 34 g 115 t 2 others

Query Match 2.7%; Score 21; DB 24; Length 284;

Best Local Similarity 78.4%; Pred. No. 7.94e-03; Mismatches 0; Indels 0; Gaps 0;

Matches 29; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

Db 182 cattaaacaatattttcttaaagtataaaatata 218

QY 371 CATTAAACAATATTCTTAAGTATAAAATATA 407

AACTGAGAATGATTAGCTTGTGTTTTTTTTTTT 3', double-stranded

cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Pac I and cloned into the Pac I and Eco RI sites of the modified pT73 vector. Library went through one round of normalization. Library constructed by Bento Soares and M.Fatima Bonaldo.

Eukaryota; Metazoa; Chordata; Vertebrata; Gnathostomata; Mammalia;

Buteraria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 294)

AUTHORS Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,

Holman,M., Hiltzman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,

Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F., Trevaskis,E., Waterston,R., Williamson,A., Wohldmann,P. and

Wilson,R. The WashU-Merck EST Project

unpublished (1995)

COMMENT Contact: Wilson RK

WashU-Merck EST Project

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: est@wustl.edu

High quality sequence stops: 271

Source: IMAGE Consortium, LBL

This clone is available royalty-free through LBL ; contact the

IMAGE Consortium (info@image.lnl.gov) for further information.

Location/Qualifiers

1..294

/organism="Homo sapiens"

/clone="116989"

BASE COUNT 97 a 62 c 49 g 85 t 1 others

ORIGIN

Query Match 2.7%; Score 21; DB 10; Length 294;

Best Local Similarity 88.9%; Pred. No. 7.94e-03; Mismatches 0; Indels 0; Gaps 0;

Matches 24; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Db 256 taatattaaatgtacttcaactgttaaga 282

QY 99 TAAATGATATTACITTAACCTAACAA 125

Search completed: Fri Dec 5 08:08:05 1997

Job time : 502 secs.

FEATURES SOURCE

ORIGIN

Query Match 2.7%; Score 21; DB 24; Length 284;

Best Local Similarity 78.4%; Pred. No. 7.94e-03; Mismatches 0; Indels 0; Gaps 0;

Matches 29; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

Db 182 cattaaacaatattttcttaaagtataaaatata 218

QY 371 CATTAAACAATATTCTTAAGTATAAAATATA 407

BASE COUNT 95 a 38 c 34 g 115 t 2 others

ORIGIN

Query Match 2.7%; Score 21; DB 24; Length 284;

Best Local Similarity 78.4%; Pred. No. 7.94e-03; Mismatches 0; Indels 0; Gaps 0;

Matches 29; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

Db 182 cattaaacaatattttcttaaagtataaaatata 218

QY 371 CATTAAACAATATTCTTAAGTATAAAATATA 407

BASE COUNT 95 a 38 c 34 g 115 t 2 others

ORIGIN

Query Match 2.7%; Score 21; DB 24; Length 284;

Best Local Similarity 78.4%; Pred. No. 7.94e-03; Mismatches 0; Indels 0; Gaps 0;

Matches 29; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

Db 182 cattaaacaatattttcttaaagtataaaatata 218

QY 371 CATTAAACAATATTCTTAAGTATAAAATATA 407

BASE COUNT 95 a 38 c 34 g 115 t 2 others

ORIGIN

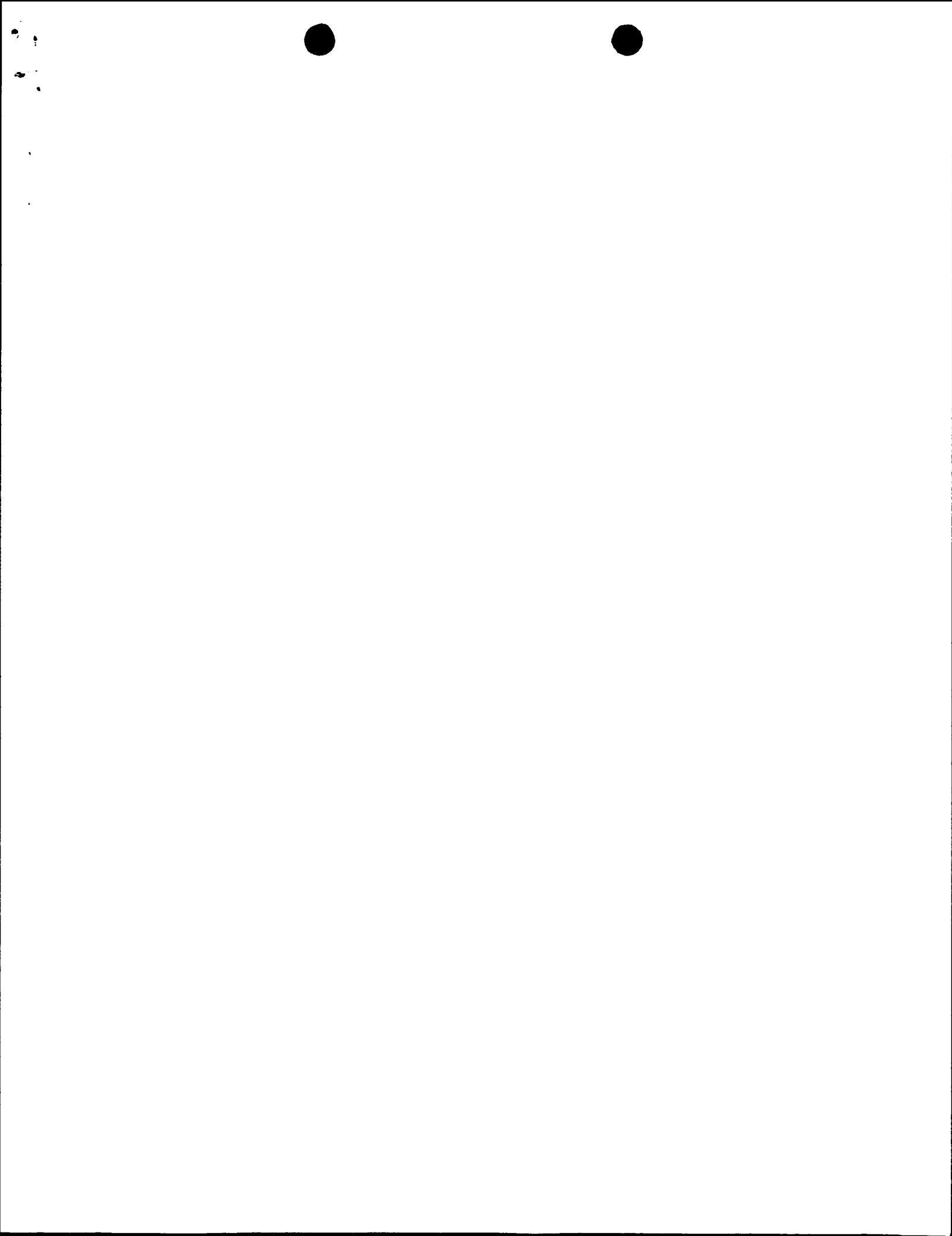
Query Match 2.7%; Score 21; DB 24; Length 284;

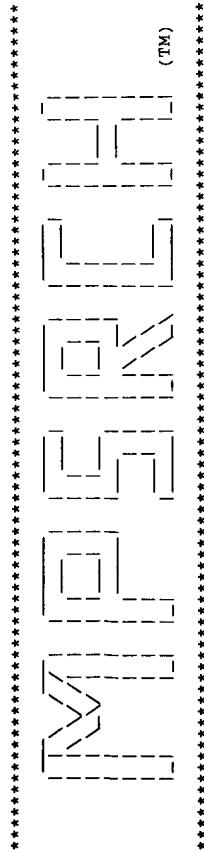
Best Local Similarity 78.4%; Pred. No. 7.94e-03; Mismatches 0; Indels 0; Gaps 0;

Matches 29; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

Db 182 cattaaacaatattttcttaaagtataaaatata 218

QY 371 CATTAAACAATATTCTTAAGTATAAAATATA 407





Release 2.1D, John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993, 1995 University of Edinburgh, U.K.  
Distribution rights by Intelligentec, Inc.

Description: (1-771) from US08790043A.seq

Perfect Score: 771

N.A. Sequence: ATGTTAAATCTGTAAAACAA.....GATTCACGCATTAATAAA 771

Comp: TACAATTTAGAACCTTGT.....CTAAGGGCGTTAATTATT

Similar output not generated.

Title: >US-08-790-043A-2

(1-771) from US08790043A.seq

Post-processing: Minimum Match 0%

Listing first 45 summaries

Scoring table: TABLE default

Gap 6

Ninmatch STD : Dbase 0; Query 0

Searched: 359085 seqs, 137405154 bases x 2

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database: EST-THREE

1:EST199 2:EST200 3:EST201 4:EST202 5:EST203 6:EST204

7:EST205 8:EST206 9:EST207 10:EST208 11:EST209 12:EST210

13:EST211 14:EST212 15:EST213 16:EST214 17:EST215

18:EST216 19:EST217 20:EST218 21:EST219 22:EST220

23:EST221 24:EST222 25:EST223 26:EST224 27:EST225

28:EST226 29:EST227 30:EST228 31:EST229 32:EST230

33:EST231 34:EST232 35:EST233 36:EST234 37:EST235

39:EST24 40:EST25 41:EST26 42:EST27 43:EST28 38:EST29

46:EST211 47:EST212 48:EST213

EST-FOUR

49:gnEST1 50:gnEST2 51:gnEST3 52:gnEST4 53:gnEST5

54:gnEST6 55:gnEST7 56:gnEST8 57:gnEST9 58:gnEST10

59:gnEST11 60:gnEST12 61:gnEST13 62:gnEST14 63:gnEST15

64:gnEST16 65:gnEST17 66:gnEST18 67:gnEST19 68:gnEST20

69:gnEST21 70:gnEST22 71:gnEST23 72:gnEST24 73:gnEST25

74:gnSTS75 75:enEST1 76:enEST2 77:enEST3 78:enEST4

79:enEST80 81:enEST6 82:enEST7 83:enEST9

84:enEST10 85:enEST11 86:enEST12 87:enEST13 88:enEST14

89:enEST15 90:enEST16 91:enEST17 92:enEST18 93:enEST19

94:enEST20 95:enEST21 96:enEST22 97:enEST23 98:enEST24

99:enEST25 100:enEST26 101:enEST27 102:enEST28

103:enEST29 104:enEST30 105:enEST31 106:enEST32 107:ueEST1

108:ueEST2

Statistics: Mean 10.599; Variance 2.102; scale 5.042

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the total score distribution, and is derived by analysis of the total score distribution.

SUMMARIES  
Result      8      Query

No.	Score	Match Length	DB	ID	Description	Pred. No.
1	23	3.0	396	25	N59624	2.18e-05
c	22	2.9	233	85	HS116643	3.39e-04
c	22	2.9	233	163	zs1006.r1	3.39e-04
c	4	2.9	393	49	AA125483	3.39e-04
c	22	2.9	393	7	AA125483	3.39e-04
c	22	2.9	424	50	mq8809.r1	3.39e-04
c	7	2.9	427	71	AA251305	3.39e-04
c	8	2.9	427	85	HS116622	3.39e-04
c	9	2.7	333	108	MM113919	4.77e-03
c	10	2.7	341	52	AA221186	4.77e-03
c	11	2.7	341	52	my25c01.r1	4.77e-03
c	12	2.7	358	13	AA185832	4.77e-03
c	13	2.7	394	13	AA185832	4.77e-03
c	14	2.7	408	30	NE7897	4.77e-03
c	15	2.7	409	95	MM490332	4.77e-03
c	16	2.7	415	27	W36178	4.77e-03
c	17	2.7	449	18	AA200846	4.77e-03
c	18	2.7	449	61	AA200846	4.77e-03
c	19	2.7	449	104	MNA8446	4.77e-03
c	20	2.1	593	19	AA059335	4.77e-03
c	21	2.6	135	55	AA088037	5.99e-02
c	22	2.6	193	36	DM96D10S	5.99e-02
c	23	2.6	196	42	GI10638	5.99e-02
c	24	2.6	250	85	HS1166217	5.99e-02
c	25	2.6	250	71	AA51371	5.99e-02
c	26	2.6	328	39	G10965	5.99e-02
c	27	2.6	348	42	G22403	5.99e-02
c	28	2.6	348	1	W8413	5.99e-02
c	29	2.6	400	41	GI17008	5.99e-02
c	30	2.6	412	21	AA019095	5.99e-02
c	31	2.6	422	64	AA230876	5.99e-02
c	32	2.6	422	99	MMA03876	5.99e-02
c	33	2.6	435	42	G21132	5.99e-02
c	34	2.6	445	51	AA169090	5.99e-02
c	35	2.6	463	2	W94669	5.99e-02
c	36	2.6	463	90	HS96694	5.99e-02
c	37	2.6	470	30	AA055104	5.99e-02
c	38	2.6	473	77	AT5074	5.99e-02
c	39	2.6	486	22	AA211548	5.99e-02
c	40	2.6	532	77	AT5070	5.99e-02
c	41	2.6	559	101	MM4901917	5.99e-02
c	42	2.6	585	9	AA131271	5.99e-02
c	43	2.6	676	34	AA11420	5.99e-02
c	44	2.6	1990	38	G06577	5.99e-02
c	45	19	288	54	AA215434	5.99e-02

## ALIGNMENTS

RESULT	1	N59624	396 bp mRNA	EST
LOCUS		Y74c01..s1	Soares fetal liver spleen 1NFLS Homo sapiens cDNA clone	
DEFINITION		Y74c01..s1		
ACCESSION		N59624		
NID		91205514		
REFERENCE		1 (bases 1 to 396)		
AUTHORS		Hillier,L., Clark,N., Duboule,T., Elliston,K., Hawkins,M., Holman,M., Hultman,M., Kucaba,T., Lennon,G., Marr,M., Parsons,J., Rifkin,L., Roblfing,T., Tan,F., Trevaskis,E., Watson,R., Williamson,A., Wohldmann,P. and Wilson,R.		
TITLE		WashU-Merck EST Project		
JOURNAL		Unpublished (1995)		
COMMENT		Contact: Wilson RK WashU-Merck EST Project Washington University School of Medicine		







FT Bento Soares and M. Fatima Bonaldo."  
 FT /clone="68475"  
 FT /clone\_lib="Soares NBINGBC"  
 FT /tissue\_type="Germina B-cell"  
 FT /lab\_host="DH10B"  
 FT mRNA Sequence 427 BP; 155 A; 69 C; 60 G; 143 T; 0 other;  
 SQ 09 Query Match 2.9%; Score 22; DB 85; Length 427;  
 Best Local Similarity 80.0%; Pred. No. 3.39e-04;  
 Matches 36; Conservative 0; Mismatches 8; Indels 1; Gaps 1;  
 DE 355 agtcacaaatgtatgttgcattcaataaaaaattaaagcaatc 398  
 QY 589 AGTGCAGAAAGGTGCGGGCTTCATACATTTAAAGAAC 633

RESULT 9 LOCUS AA072168 333 bp mRNA EST  
 DEFINITION mm69c05.r1 Stratogene mouse macrophage (#37306) Mus musculus CDNA  
 VERSION AA072168 5; similar to gb:X65873 KINESIN HEAVY CHAIN (HUMAN);  
 g1590513  
 KEYWORDS EST.  
 SOURCE house mouse.  
 ORGANISM Mus musculus; mitochondrial eukaryotes; Metazoa; Chordata;  
 Vertebrata; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
 Mus.

COMMENT Contact: Marra M/Mouse EST Project  
 Marra M., Hillier L., Allen M., Bowles M., Dietrich N., Dubuque T.,  
 Morris J., Le M., Martin J., Lacy M., Moore B., Underwood K.,  
 Schellenberg K., Steptoe M., Tan F., Underwood K.,  
 Theising B., Wylie T., Lennon G., Wilson R., and  
 Waterston R.  
 The WashU-HMMI Mouse EST Project  
 Unpublished (1996)

REFERENCE Marra M., Hillier L., Allen M., Bowles M., Dietrich N., Dubuque T.,  
 Morris J., Le M., Martin J., Lacy M., Moore B., Underwood K.,  
 Schellenberg K., Steptoe M., Tan F., Underwood K.,  
 Theising B., Wylie T., Lennon G., Wilson R., and  
 Waterston R.

AUTHORS Tel: 314 286 1800  
 Fax: 314 886 110  
 Email: mouse@watson.wustl.edu

KEYWORDS This clone is available royalty-free through LLNL; contact the  
 IMAGE Consortium (info@image.llnl.gov) for further information.  
 MGI:320004 Seq primer: -28m13 rev1 ET from Amersham  
 High quality sequence stop: 91.

JOURNAL Location/Qualifiers  
 FEATURES source 1..333  
 /organism="Mus musculus"  
 /note="Vector: pBluescript SK-; Site\_1: EcoRI; Site\_2:  
 XbaI; Cloned unidirectionally. Primer: Oligo dt. WEHI-3  
 cell line. Average insert size: 1.5 kb; Uni-ZAP XR vector;  
 ~5' adaptor sequence: 5' GAATTCCGACGAG 3' -3'; adaptor  
 sequence: 5' CTCTGAGTTTCTTCTTCTTCTT 3'."  
 /clone="533668"  
 /dev\_stage="Stratagene mouse macrophage (#937306)"  
 /lab\_host="SOLR (kanamycin resistant)"  
 <1..>333

BASE COUNT 101 a 79 c 93 g 60 t

Query Match 2.7%; Score 21; DB 32; Length 333;  
 Best Local Similarity 61.8%; Pred. No. 4.77e-03;  
 Matches 55; Conservative 0; Mismatches 34; Indels 0; Gaps 0;

Db 51 tgccagaaggatgaaaatgtttggatcgccgtggaaaggccaaagaaaatgtatctcg 110  
 Qy 591 TGCAGAAAAGGTGCGGGCTTCATACATTTAAAGAAATCGAGGGTGCACCTT 650

RESULT 10 ID MM139719 standard; RNA; EST; 341 BP.  
 AC AA221186;  
 NI 91840172  
 DT 14-FEB-1997 (Rel. 50, Created)  
 DT 14-FEB-1997 (Rel. 50, Last updated, Version 1)  
 DE my25c01.r1 Baustead mouse pooled organs MPLRB4 Mus musculus cDNA  
 clone 696664 5'.  
 KW EST.  
 OS Mus musculus (house mouse)  
 OC Eukaryote; mitochondrial eukaryotes; Metazoa; Chordata;  
 Vertebrata; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
 OC Mus.  
 RN [1]  
 RP 1-341  
 RA Marra M., Hillier L., Allen M., Bowles M., Dietrich N.,  
 Dubuque T., Geisel S., Kucaba T., Lacy M., Le M., Martin J.,  
 Morris M., Schellenberg K., Steptoe M., Tan F., Underwood K.,  
 Moore B., Theising B., Wylie T., Lennon G., Wilson R.,  
 Waterston R.;  
 RA "The WashU-HMMI Mouse EST Project";  
 RT Unpublished.  
 RL CC Washington University School of Medicinep 4444 Forest Park Parkway,  
 Box 8501, St. Louis, MO 63108 Tel: 314 286 1800 Fax: 314 286 1810  
 CC Email: mouse@watson.wustl.edu This clone is available  
 royalty-free through LLNL; contact the IMAGE Consortium  
 (info@image.llnl.gov) for further information. MGI:430424 Seq  
 CC Primer: -28m13 rev1 ET from Amersham High quality sequence stop:  
 CC 234.  
 CC Key  
 FH Location/Qualifiers  
 FH source 1..341  
 FT /organism="Mus musculus"  
 FT /strain="FVB/N"  
 FT /note="Vector: pT7T3D-Pac (Pharmacia) with a modified  
 polylinker; Site\_1: EcoRI; Site\_2: NotI; 1st strand cDNA  
 was primed with a Not I - Oligo (dt) primer [5'-  
 TGTTACGATCTGAGTGGAGGGCGGCCCTTTTTTTTTTTTTTTTTTTTTTTTTTT  
 3']'; double-stranded cDNA was ligated to Eco RI adaptors  
 [GTTGATTCGGTACCC], digested with Not I and cloned into the  
 IMAGE Consortium (info@image.llnl.gov) for further information.  
 Not I and Eco RI sites of the modified pT7T3 vector.  
 Library constructed by Bob Barstead."  
 /clone="696864"  
 /clone\_id="Barstead mouse pooled organs MPLRB4"  
 /sex="mixed"  
 /dev\_stage="7 day"  
 /lab\_host="DH10B"  
 FT mRNA Sequence 341 BP; 101 A; 67 C; 73 G; 100 T; 0 other;  
 SQ Query Match 2.7%; Score 21; DB 108; Length 341;  
 FT Best Local Similarity 81.8%; Pred. No. 4.77e-03;  
 FT Matches 27; Conservative 0; Mismatches 6; Indels 0; Gaps 0;  
 Db 269 acaatgttaatgaaaccataaaaatgttcttg 301  
 CP 383 ACAATGTATGTTAGTAAGAACACTATGTCCTGA 351

RESULT 11 LOCUS AA221186 341 bp mRNA  
 DEFINITION my25c01.r1 Barstead mouse pooled organs MPLRB4 Mus musculus cDNA  
 clone 696664 5'.  
 ACCESSION AA221186  
 NID 91840172  
 KEYWORDS EST.

SOURCE	house mouse.	TITLE	Current status of the Plasmodium falciparum genome project
ORGANISM	Mus musculus	JOURNAL	Mol. Biochem. Parasitol. 79, 1-12 (1996)
REFERENCE	1 (bases 1 to 341)	COMMENT	97/001675
AUTHORS	Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T., Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M., Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B., Waterston, R., Wylie, T., Lennon, G., Soares, B., Wilson, R. and The WashU-HMI Mouse EST Project	FEATURES	Contact: Debopam Chakrabarti Department of Molecular Biology and Microbiology University of Central Florida Orlando, FL 32816-2360 Tel: 407 384 2061 Fax: 407 384 3095 Email: dchakrap@pegasus.cc.ucf.edu Seq primer: T3. Location/Qualifiers
TITLE	Unpublished (1996)	source	1..358 <1..>358 /organism="Plasmodium falciparum" /strain="Dd2"
JOURNAL		FEATURES	/note="vector: Lambda ZAP II; Site_1: Ecor I; Site_2: Xba I; PolyA+ RNA, from asynchronous blood stage parasites of the Dd2 isolate cultured in vitro, was reverse transcribed using an oligo dt-Xba I primer. Second strand was prepared using RNase H and DNA Polymerase I. EcoRI adaptors were ligated to the cDNA, and it was digested with Xba I. Prepared fragments were ligated into Ecor I + Xba I digested lambda ZAP II vector." 'clone'_lib="PF1470C" 'clone'_lab="czappFDd2.1", Debopam Chakrabarti" /lab_host="E. coli XL-1 blue"
COMMENT	Contact: Marra M/Mouse EST Project WashU-HMI Mouse EST Project Washington University School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108 Tel: 314 286 1800 Fax: 314 286 1810 Email: mouseest@watson.wustl.edu This clone is available royalty-free through LInL; contact the IMAGE Consortium (info@image.llnl.gov) for further information. MGI: 1430424 Seq primer: -28m13 rev2 ET from Amersham High quality sequence stop: 234.	source	BASE COUNT mRNA ORIGIN Query Match 13 AA185532 394 bp mRNA EST Best Local Similarity 2.7% Score 21; DB 1; Length 358; Matches 27; Conservative 0; Mismatches 4; Indels 0; Gaps 0; ACCESSION AA185532 KEYWORDS EST. SOURCE blood fluke. ORGANISM Schistosoma mansoni DEFINITION Schistosomatidae; Schistosoma mansoni, adult worm (Pena,H.B. and Rabelo,E.M.L., 1997) AUTHORS Pena,S.D.J.) Schistosoma mansoni cDNA clone SMA0115 5' end. TITLE Analysis of cDNA libraries from different developmental stages of Schistosoma mansoni with the aim of producing Expressed Sequence Tags JOURNAL Unpublished (1996)
SOURCE	"Mus musculus"	FEATURES	RESULT 13 AA185532 394 bp mRNA EST LOCUS 08-JAN-1997 DEFINITION MAAD0115.M2R Schistosoma mansoni, adult worm (Pena,H.B. and Rabelo,E.M.L., 1997) ORGANISM Schistosomatidae; Schistosoma mansoni ACCESSION AA185532 NID 91769646 KEYWORDS EST. SOURCE blood fluke. ORGANISM Schistosoma mansoni DEFINITION Schistosomatidae; Schistosoma mansoni AUTHORS Rabelo,E.M.L., Franco,G.R., Azevedo,V., Pena,H.B., Santos,T.M., Neira,W.S.P., Rodrigues,N.A., Ortega,J.M. and Pena,S.D.J., TITLE Analysis of cDNA libraries from different developmental stages of Schistosoma mansoni with the aim of producing Expressed Sequence Tags JOURNAL Unpublished (1996)
ORGANISM	"Mus musculus"	FEATURES	RESULT 13 AA185532 394 bp mRNA EST LOCUS 08-JAN-1997 DEFINITION MAAD0115.M2R Schistosoma mansoni, adult worm (Pena,H.B. and Rabelo,E.M.L., 1997) ORGANISM Schistosomatidae; Schistosoma mansoni ACCESSION AA185532 NID 91769646 KEYWORDS EST. SOURCE blood fluke. ORGANISM Schistosoma mansoni DEFINITION Schistosomatidae; Schistosoma mansoni AUTHORS Rabelo,E.M.L., Franco,G.R., Azevedo,V., Pena,H.B., Santos,T.M., Neira,W.S.P., Rodrigues,N.A., Ortega,J.M. and Pena,S.D.J., TITLE Analysis of cDNA libraries from different developmental stages of Schistosoma mansoni with the aim of producing Expressed Sequence Tags JOURNAL Unpublished (1996)
REFERENCE	1 (bases 1 to 358)	FEATURES	Contact: Pena, H. B. and Pena, S. D. J. Laboratorio de Genetica-Bioquimica, Departamento de Bioquimica e Imunologia Instituto de Ciencias Biologicas, Universidade Federal de Minas Gerais Avenida Antonio Carlos 6627, Belo Horizonte, MG, Brazil, 31270-010
AUTHORS	Coppel, R., Cowman, A., Craig, A., Chakrabarti, D., Christodoulou, Z., Dame, J. B., Arnott, D. E., Bourke, P., Fischer, K., Foster, J. J., Goodman, N., Hinterberg, K., Holder, A. A., Holt, D., Kemp, D., Lanzer, M., Lim, A., Newbold, C., Ravetch, J. V., Reddy, G. R., Rubio, J., Schuster, S. E., Su, X., Thompson, J. R., Vital, F., Wellens, T. E. and Werner, E.	FEATURES	1..394 /organism="Schistosoma mansoni"



Db 285 aaaaatgatgaagtgcgtggattttggcggaaaaggccatgtcttaggtatattc 344  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
Qy 203 AAAAGCGATGAGGGTTATTAAATGGTTTGCCAAATTGTTAAAGATGTGGCAATATTG 262  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
Db 345 ctggagttt 351  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
Qy 263 ATGGTGT 269  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

Search completed: Fri Dec 5 08:12:53 1997  
Job time : 266 secs.